Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
HERSCHEL W. LEIBOWITZ, *International Business Machines Corporation, Bethesda, Maryland*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Minimal Paired-Associate Learning: L. R. PETERSON AND M. J. PETERSON 521
Studies in Concept Attainment: III. Effect of Instructions at Two Levels of Intelligence: S. F. OSLER AND S. R. WEISS 528
Inhibitory Potential in Rotary Pursuit Acquisition by Normal and Defective Subjects: R. W. JONES AND N. R. ELLIS 534
The Generality of the "Conditioning Axiom" in Human Binary Prediction: P. L. DERKS 538
Multidimensional Scaling of Facial Expressions: R. P. ABELSON AND V. SERMAT 546
Spontaneous Recovery Interval as a Factor in Reacquisition of T Maze Behavior: J. W. COTTON, G. D. JENSEN, AND D. J. LEWIS 555
Operant Conditioning of the GSR: R. L. FOWLER AND H. D. KIMMEL 563
An Analysis of Set in Relation to Reaction Time: J. BOTWINICK AND J. F. BRINLEY 568
Donders' B- and C-Reactions and S-R Compatibility: D. E. BROADBENT AND M. GREGORY ... 575
Taste and Consummatory Activity in Amount and Gradient of Reinforcement Functions: F. A. KNARR AND G. COLLIER 579
Activity during Delay of Reinforcement in Human Learning: R. A. CHAMPION AND D. A. MCBRIDE 589
The Distribution of Recalled Items in Simultaneous Intentional and Incidental Learning: A. MECHANIC 593
Consistency and Complexity of Response Sequences as a Function of Schedules of Noncontingent Reward: J. C. WRIGHT 601
Saccharine Concentration and Deprivation as Determinants of Instrumental and Consummatory Response Strengths: H. L. SNYDER 610
Concept Identification as a Function of Obviousness of Relevant and Irrelevant Information: E. J. ARCHER 616
Reciprocal Augmentation of Generalization and Anxiety: S. A. MEDNICK AND C. WILD 621

American Psychological Association

Vol. 63, No. 6 June 1962

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.* When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1962 the publication cost is $30 per printed page.

All **regular** articles, **early publication** articles, **supplementary reports,** and **replication reports** should be submitted to the Editor-elect: David A. Grant: Department of Psychology, University of Wisconsin; Madison 6, Wisconsin. Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts.* Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR
Acting Production Manager

RUTH P. JAGOLINZER
Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1962 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 63, NO. 6 JUNE 1962

MINIMAL PAIRED-ASSOCIATE LEARNING [1]

LLOYD R. PETERSON AND MARGARET JEAN PETERSON

Indiana University

Paired-associate learning can be conceptualized as complex discrimination learning. The simplest analysis should be possible where the number of pairs is reduced to a minimum of two, in which case the parallel to simple discrimination learning is apparent. Presumably paired-associate learning in this limiting case has not been investigated because it is too easy, and therefore a proper learning curve could not be obtained. However, it has been shown that an even simpler phenomenon, recall of single items, can be studied effectively by using retention over short intervals as the measure of response strength (Peterson & Peterson, 1959; Peterson, Peterson, & Miller, 1961). The extension of this technique to paired associates requires the addition of a cue for the response as well as a second response and its cue. The consideration that a specific cue is provided for the response might be expected to improve retention, but with the introduction of a second response the difficulty of the task should be increased. The series of experiments to be reported investigated the interference effects from paired associates presented either just before or just after the pair to be tested.

[1] This research was supported by Grant G12917 from the National Science Foundation to Indiana University for which the senior author is principal investigator.

GENERAL METHOD

Subjects.—The *Ss* were students from introductory psychology courses at Indiana University. Participation in experiments was a course requirement.

Apparatus.—A memory drum was mounted in a 3 × 4 ft. black screen. A sliding aperture permitted exposure of seven columns each containing 28 spaces. The drum motor was controlled by two decade interval timers.

Test materials.—The materials were pairs of English words having frequencies of occurrence of at least 10 per million (Thorndike & Lorge, 1944). The first member of a pair was a three-letter word typed in red capital letters, and the second was a four-letter word typed in black lower case letters. The words were paired randomly with the restriction that pairings in which the words were obviously related were eliminated. Pairs tested for recall appeared equally often in each condition of an experiment over a group of *Ss*. A given pair was presented only once to an individual *S*.

Procedure.—The *Ss* were instructed to look at the stimulus words and to pronounce aloud the black four-letter response words in Exp. I. In Exp. II through IV *Ss* were told to pronounce aloud both stimulus and response words. The *Ss* were instructed that each

TABLE 1

ANALYSIS OF VARIANCE OF MEAN PROPORTIONS CORRECT IN EXP. I

Source	*df*	*MS*	*F*
Presentation (P)	2	3.383	26.02**
Retention (R)	2	.237	8.46**
Ss	26	.110	
P × R	4	.014	.48
P × *Ss*	52	.130	4.48**
R × *Ss*	52	.028	.97
P × R × *Ss*	104	.029	

** $P < .01$.

stimulus word was paired with a response word and that the pair was to be learned. Words were exposed one at a time for about 1 sec. each. A stimulus word appeared followed by a response. Depending on the condition one or two other pairs might follow. A randomly selected three-digit number then appeared, following which the drum stopped on a blank space. The *Ss* counted backward by 2, 3, or 4 as rapidly as possible when the number appeared and continued until the signal for recall. Instructions for a change in counting were given during 1-min. rest periods coincident with a change of the tape. The signal for recall was a red light in Exp. I and a bell in Exp. II through IV. Accompanying this signal was an exposure of one of the stimulus words, and *S* was instructed to recall the response word that had been previously paired with this stimulus. The retention interval was measured from removal of the response to be tested from the aperture to appearance of the cue for recall. Fifteen seconds after the signal for recall in Exp. I, 12 sec. in Exp. II through IV, a new cycle was begun by the flashing of a green light as a blank space was exposed. There followed a new stimulus term. Practice tests were given following the instructions, in which *Ss* were corrected for procedural errors and told to count as fast as possible if they hesitated.

EXPERIMENT I

Method

Each of 27 *Ss* was tested six times in every condition of a 3 × 3 factorial design consisting of three conditions of presentation tested at retention intervals of 4, 8, and 16 sec. In Cond. S (Single) only one pair of words was presented. In Cond. DA (Double A) two pairs were presented and the first was tested. In Cond. DB (Double B) two pairs were presented and the second was tested. Each condition appeared once per block of nine tests in a random position. The *Ss* pronounced the response but not the stimulus. Six practice tests were given.

Results and Discussion

The analysis of variance summarized in Table 1 indicates that both retention interval and presentation condition influenced correct recalls. Recall in the two-pair conditions (Table 2) was markedly lower than that in the one-pair conditions, as would be predicted from considerations of interference. The superiority of DA over DB is a complicating circumstance, although the Presentation × *Ss* interaction appears to indicate an absence of the majority trend with some of the *Ss*.

Errors are classified as to source in Table 2. The term "alternate" refers

TABLE 2

MEAN PROPORTIONS RECALLED IN EXP. I

Measure	Cond. S Intervals			Cond. DA Intervals			Cond. DB Intervals		
	4	8	16	4	8	16	4	8	16
Correct	.94	.89	.84	.67	.63	.60	.57	.46	.43
Alternates				.27	.22	.26	.35	.39	.36
Stimuli	.01	.00	.01	.00	.01	.01	.02	.02	.02
Previous	.01	.04	.04	.01	.06	.05	.03	.04	.06
Extraneous	.03	.00	.00	.01	.02	.01	.01	.00	.03
Omissions	.01	.07	.11	.04	.06	.07	.02	.09	.10

to the nontested response member of the current cycle, and "stimuli" are those from that same cycle. "Previous" errors include stimuli and responses from earlier cycles. "Extraneous" errors include any overt response not previously used in the experiment. It is apparent that recall of the alternate response accounts for the bulk of the errors of commission. The greater incidence of alternate responses presented first as compared to those presented second is reversed from what would be predicted on the basis of recency, for the time since presentation of this competing response is longer in Cond. DB than in Cond. DA.

An explanation of the primacy effect might be that *Ss* were required to read aloud only the responses, and thus they may have failed to reliably observe the second stimulus. A test of this hypothesis was made by requiring *Ss* to read aloud both stimuli and responses. In addition, the number of pairings was varied to ascertain if repetition might influence the primacy effect.

Experiment II

Method

A 2 × 2 within-*Ss* design was used with 24 *Ss*. A set of two pairs was presented either once or twice, and either the first or second pair was tested. In the repetition conditions three orders of presentation occurred equally often for each *S*: AABB, ABAB, and ABBA. Each of the four main conditions appeared equally often in a random position in each block of 12 tests. Each condition was tested 12 times following 4 practice tests. The retention interval was 8 sec., and the intercycle interval was 12 sec.

TABLE 3

Analysis of Variance of Mean Proportions Correct in Exp. II

Source	*df*	*MS*	*F*
Repetition (R)	1	2.083	40.84**
Order (O)	1	.631	7.09*
Half (H)	1	.148	2.51
Ss	23	.163	
R × O	1	.020	
R × H	1	.005	
R × *Ss*	23	.051	1.34
O × H	1	.020	
O × *Ss*	23	.089	2.34**
H × *Ss*	23	.059	1.55
R × O × H	1	.099	2.61
Error (res.)	92	.038	

* $P < .05$.
** $P < .01$.

Results and Discussion

The data were analyzed in such a way that first and second halves of the session could be compared, but the analysis of variance (Table 3) did not show this to be a significant variable. The primacy effect persisted with repetition, as can be seen in Table 4,

TABLE 4

Mean Proportions Recalled in Exp. II

Measure	First Half				Second Half			
	One Rep.		Two Rep.		One Rep.		Two Rep.	
	A	B	A	B	A	B	A	B
Correct	.40	.38	.69	.53	.53	.38	.71	.60
Alternates	.14	.12	.12	.20	.08	.17	.13	.19
Stimuli	.06	.05	.06	.04	.05	.08	.06	.04
Previous	.12	.08	.03	.05	.06	.07	.02	.01
Extraneous	.02	.04	.02	.01	.03	.02	.04	.03
Omissions	.26	.33	.08	.17	.25	.28	.04	.13

the repetition increasing recall for both first and second pairs. The *Ss* rarely made errors in reading, and no errors involved the second pair, so there can be no question but that *Ss* observed the second response and its cue. The primacy effect persisted with all sequences of presentation in the repetition conditions. Proportions correct when the first pair was tested were .71, .68, and .70 for the AABB, ABAB, and ABBA sequences, respectively. The proportions in testing of the second pair were .55, .57, and .58, in the same order.

Recall in the one-presentation conditions was lower than in Exp. I at the 8-sec. retention interval. This decrease may be related to changes in error frequencies seen by comparing Tables 2 and 4. Occurrences of alternate responses were sharply reduced in Exp. II, while stimulus errors increased somewhat. These changes may be related to the fact that *S* read the stimuli aloud in Exp. II. Frequency of words from previous cycles increased somewhat in Exp. II, probably reflecting vocalization of stimuli, repeated presentations, and a decrease in time between cycles. Errors from previous cycles decreased slightly in the second half of the session, and in either session accounted for only a small proportion of the errors. The largest change from Exp. I to Exp. II was an increase in the number of tests in which no response was given, indicating that the overt errors provide only a partial analysis of the forgetting.

TABLE 5

ANALYSIS OF VARIANCE OF MEAN PROPORTIONS CORRECT IN EXP. III

Source	*df*	*MS*	*F*
Order (O)	2	.227	8.74***
Half (H)	1	.181	6.46*
Ss	35	.154	7.33
O × H	2	.016	.76
O × *Ss*	70	.026	1.24
H × *Ss*	35	.028	1.33
O × H × *Ss*	70	.021	

* $P < .05$.
*** $P < .001$.

TABLE 6

MEAN PROPORTIONS RECALLED IN EXP. III

Measure	First Half			Second Half		
	A	B	C	A	B	C
Correct	.28	.24	.18	.36	.26	.25
Alternates	.13	.18	.18	.18	.16	.15
Stimuli	.09	.07	.05	.07	.05	.08
Previous	.08	.09	.09	.05	.06	.08
Extraneous	.02	.07	.05	.02	.03	.03
Omissions	.40	.35	.45	.32	.44	.41

EXPERIMENT III

The persistence of the primacy effect in Exp. II led to an investigation of recall with three pairs of associates. It was considered possible that the middle position might prove to be inferior in recall to both first and last positions as in serial list learning and immediate memory experiments (Broadbent, 1958).

Method

Three pairs were presented once, and after a retention interval of 12 sec., the bell sounded at the same time that one of the three stimuli appeared to cue recall. Each condition was tested 16 times with each of 36 *Ss*. The sequence of conditions was randomized within each of four blocks of 12 tests. The interval from recall cue to next cycle was 12 sec. Four practice cycles were given.

Results and Discussion

The analysis of variance summarized in Table 5 shows that again order in presentation was a significant variable. However, no bowing of the position curve was found (Table 6). There was a straight primacy effect.

There was a significant increase in recall in the second half of the session, but this did not interact with presentation order.

The alternate errors tabulated in Table 5 include both of the other response members within the cycle being tested, since no differences between them appeared reliable. Again, errors from the current cycle predominated, and again there was a slight decrease in errors from previous cycles during the second part of the session.

Experiment IV

The previous experiments showed a decrement in recall related to the number of pairs within a cycle. Response interference seemed a likely explanatory factor, but it did not account for the persistent serial position effect. A factor influencing recall which might produce a primacy effect is stimulus generalization decrement. If the stimulus situation for the response presented last were to be considered to extend in time to include the preceding pair of words, then the stimulus at recall differs in that the two preceding words are now absent. Stimulus conditions at recall for the pair presented first are more similar to those of presentation than for the second pair, since no words precede its cue in either presentation or recall. Thus, superiority of recall would be predicted for the first pair by reason of generalization decrement. An experiment was designed which increased the number of cues at recall by presenting the alternate pair as part of the stimulus situation.

Method

Fifty-six *Ss* were assigned in order of appearance to a randomized list partitioning them into two groups. Group S (Silent) was instructed to look at the words which appeared at recall but not to speak them aloud. Group V (Vocal) was instructed to speak each word aloud. Both groups read the words in the original presentation aloud. Two pairs of words were presented in each cycle. Each *S* served in all conditions of a 2×2 factorial design, in which one factor was the serial position during presentation, and the other factor was the presence or absence of members of the other pair as cues for recall. For half of the 48 tests three of the previously presented words appeared one by one in the drum aperture in the order: stimulus, response, stimulus. When the drum stopped at a blank, *S* attempted to supply the correct response to the last stimulus. On the remaining tests, when the bell rang the drum showed two successive blanks, the single stimulus for the response tested appeared, and the drum stopped on a blank. The interval to beginning of the next cycle was 12 sec. Four practice cycles were given. The retention interval was 12 sec.

Results and Discussion

The analysis of variance summarized in Table 7 shows that there was, as predicted, a cue effect in recall which interacted with order in

TABLE 7

Analysis of Variance of Mean Proportions Correct in Exp. IV

Source	*df*	*MS*	*F*
Between			
Vocalization (V)	1	.010	.05
Error (b)	55	11.319	
Within			
Order (O)	1	1.856	35.02***
Cues (C)	1	.126	3.15
Half (H)	1	1.568	39.20***
O × C	1	.536	15.76***
C × H	1	.003	.09
O × H	1	.033	.97
O × V	1	.007	.21
C × V	1	.007	.21
H × V	1	.022	.65
O × C × H	1	.000	
O × C × V	1	.052	1.53
O × H × V	1	.052	1.53
C × H × V	1	.002	.06
O × C × H × V	1	.008	.24
Error (O)	55	.053	
Error (C)	55	.040	
Error (H)	55	.040	
Error (res.)	212	.034	

*** $P < .001$.

TABLE 8

MEAN PROPORTIONS RECALLED IN EXP. IV

Measure	First Half				Second Half			
	Pair A		Pair B		Pair A		Pair B	
	One Cue	Three Cues	One Cue	Three Cues	One Cue	Three Cues	One Cue	Three Cues
Group S (Silent)								
Correct	.49	.51	.28	.37	.58	.55	.42	.51
Alternates	.08	.05	.14	.07	.15	.07	.18	.03
Stimuli	.05	.01	.06	.04	.04	.02	.07	.03
Previous	.08	.05	.03	.09	.06	.05	.06	.09
Extraneous	.04	.05	.05	.06	.02	.04	.01	.06
Omissions	.26	.33	.44	.37	.15	.27	.26	.28
Group V (Vocal)								
Correct	.48	.41	.27	.39	.61	.55	.40	.51
Alternates	.09	.00	.16	.03	.10	.01	.18	.04
Stimuli	.05	.01	.06	.01	.03	.00	.10	.00
Previous	.08	.14	.07	.10	.06	.07	.05	.06
Extraneous	.02	.04	.02	.05	.02	.02	.03	.04
Omissions	.28	.40	.42	.42	.18	.35	.24	.35

presentation. Table 8 indicates that the additional cues increased recall for the pair presented second, but had no consistent influence on pairs presented first. Inspection of the error analysis indicates that occurrence of alternate responses and stimuli from the current cycle decreased under the additional cues for both positions in presentation. The increase in omissions for the A conditions with additional cues suggests that under these conditions *Ss* were able to refrain from speaking certain errors, but were not able to replace these errors with the correct response in the 12 sec. before the next cycle began. Although the hypothesis concerning the effect on serial order of stimulus generalization was supported by the data, the persisting small primacy effect suggests that other factors are also present.

The series of experiments taken together support an interference interpretation of the forgetting which occurs over short intervals of time. Little forgetting occurred with one response in a cycle, and hence proactive interference from items presented in earlier cycles does not seem of major importance. Such interference from earlier items as was evidenced by the analysis of errors, did not increase from the first half to the second half of the session. On the other hand, pairs presented either immediately before or immediately after presentation of the tested pair did markedly decrease retention. The data suggest that recency is a major variable in the relationship between previous learning and new learning.

SUMMARY

The short-term retention of minimal numbers of paired associates was studied in four experiments. Pairs of words were presented in a memory drum and tested for recall by presenting a stimulus after some interval no longer than 16 sec. during which *Ss*

counted backward. Recall decreased as number of pairs increased, but a primacy effect complicated an interpretation in terms of competing responses. This primacy effect was present with either one or two presentations of a pair. When the stimulus situation at recall was extended to include the alternate pair of words, the primacy effect was reduced. Errors at recall were largely words from the set of pairs just presented.

REFERENCES

BROADBENT, D. E. *Perception and communication*, New York: Pergamon, 1958.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

PETERSON, L. R., PETERSON, M. J., & MILLER, A. Short-term retention and meaningfulness. *Canad. J. Psychol.*, 1961, **15**, 143–147.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

(Received May 1, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 528–533

STUDIES IN CONCEPT ATTAINMENT: III. EFFECT OF INSTRUCTIONS AT TWO LEVELS OF INTELLIGENCE [1]

SONIA F. OSLER

Goucher College

AND SANDRA RAYNES WEISS

Department of Pediatrics, Johns Hopkins University

In a recent study on the effect of intelligence on concept attainment using instructions which were nonspecific as to the nature of the task, Osler and Fivel (1961) reported a significant difference in errors between groups of average and superior IQ. An important question that has yet to be answered concerns the pertinent attribute of high intelligence which accounts for the superiority of performance. One possibility is that superior *S*s supplement *E*'s instructions with their own, directing them to search for consistencies in the reinforced stimuli; whereas the less intelligent *S*s work along without self-instructions until the reinforcement contingencies of the experiment strengthen the response to the concept exemplars. Perhaps a simpler statement of this hypothesis might be that under the conditions of nonspecific instructions the task presented to *S* has two components, namely, problem *finding* as well as problem *solving*; and that the superior performance associated with high intelligence may be the result of greater proficiency at either or both of these components. On the other hand, when instructions are specific, defining the problem which *S* is expected to solve, the problem finding component of the task is eliminated or reduced to a minimum.

It was the aim of the present investigation to test these ideas experimentally. Two experiments were performed. Experiment I was a replication of the Osler and Fivel (1961) experiment. This was undertaken for the double purpose of confirming the earlier data and controlling for the possibility that automated equipment presently being used for the conduct of the experiment may have affected the results obtained. In Exp. II the general in-

[1] This investigation was supported by a grant from the Psychopharmacology Service Center, National Institute of Mental Health, made to Leon Eisenberg, Johns Hopkins University, and the senior author (Grant MY-2583).

The writers wish to express their gratitude to Mary A. Adams, Assistant Superintendent, Elementary Education; Vernon S. Vavrina, Assistant Superintendent, Secondary, Vocational, and Adult Education; and to Arthur Lichtenstein, Director of Special Services for Pupils, Baltimore City Department of Education; and to B. Melvin Cole, Director, Elementary Education; Norris A. King, Director, Secondary Education; and Charles J. Leiman, Supervisor, Psychological Services; and to Loyal W. Joos, Director, Educational Research and Planning, Baltimore County Department of Education; and to the principals and their staffs who cooperated in the conduct of the study.

The authors also wish to thank Roy Lachman, Sandra L. Shapiro, and Grace E. Trautman for testing *S*s and for many suggestions concerning the conduct of the study. To Margaret P. Martin, Associate Professor of Biostatistics, Johns Hopkins School of Hygiene and Public Health, go our special thanks for her invaluable assistance with the statistical problems encountered in this investigation.

structions were replaced by more specific ones, defining the problem for *S* and directing him to search for hypotheses.

METHOD

As Exp. I and II differed only in the instructions used, the method for both will be described together, and wherever necessary information pertinent to each experiment will be provided separately.

Subjects.—The *Ss* consisted of 180 school children attending elementary and junior high schools of Baltimore City and County, equally divided among the following age groups: 6, 10, and 14 yr. Each *S* was within 6 mo. of the age specified. Half of the *Ss* within each age level had a WISC IQ between 90 and 109 (Average Intelligence group), and the other half had a WISC IQ of 110 and above (High Intelligence group). The mean IQs of the two groups in Exp. I were 100.8 and 120.9, and in Exp. II they were 101.2 and 121.7, respectively.

Apparatus.—The apparatus has been described elsewhere (Osler & Powell, 1960). Essentially, it consisted of a film strip projector and a screen to receive the projected images of the stimuli which appeared in pairs. A telegraph key placed under each stimulus released a marble following a correct response, which consisted of pressing the key under the concept exemplar. No reinforcement followed if the key under the negative stimulus was pressed. The *Ss* were not paced, but 4 sec. after a response was made the next stimulus was presented.

The concept studied was *bird*. This was one of the three concepts on which data were reported by Osler and Fivel (1961). The stimuli used were the same as in the previous investigation, except that they now appeared on a screen projected from a film strip. The complete set of stimuli consisted of 150 pairs of pictures, one member of each pair representing a positive instance of the concept and the other a negative instance. The pictures of each pair were matched approximately in area, color, and brightness. Reasons for using naturalistic stimuli, which carry with them the hazard of unspecified dimensions, have been discussed elsewhere (Osler & Fivel, 1961). An additional reason prompting their use in this investigation was the previous finding that performance on this concept was a function of intelligence, thus demonstrating the suitability of this problem for the present study. The stimuli had been pretested on a group of children of the same age and intelligence levels as the present *Ss*, and were found to be familiar to the type of subject studied.

Procedure.—Each *S* was tested individually in two separate sessions. At the first session *S* was told that he would play a game in which he could win any one of a group of toys displayed before him. He was asked to select the most preferred of the group, and this was then set slightly apart from the others in a position which allowed *S* to view it during the "game." The *Ss* in Exp. I received the following instructions:

> This is the game we are going to play. Here are two handles (*E* points to them). If you push the correct handle you will get a marble. Sometimes this will be the correct handle and sometimes this will be the correct handle. If you push the correct handle you will get a marble like this (demonstrated). If you push the wrong handle, like this, no marble will come (demonstrated). If you get a marble you can put it here (here *E* points to a grooved wooden tray with which *S* was provided). Now you try this one and let's see what happens. If you look at these pictures before pushing the handle, you will be able to get many marbles. When you fill the tray with marbles, you can have the (name of the toy).

The stimuli used during the demonstrations were unrelated to the concept.

In Exp. II the *Ss* were also informed about a game in which they could win a toy, and then they received the following instructions:

> This is the game we're going to play. Listen carefully and I will tell you how to play the game. Here are two pictures. You pick one of them and push the handle under it, and if you are right, you will get a marble, like this (demonstrated). If you are wrong you will not get a marble, like this (demonstrated). Now you try and push this one. The game is to see how soon you can get a marble every time you choose. If you look at the pictures carefully you will see that there is something in the pictures like an idea that will tell you which one to choose to get a marble every time. Now we will start the real game. What are you going to do first? (if *S* does not say: look at the pictures, *E* supplies the answer). If you look carefully and see which pictures give you the marble you will be able to win marbles every time. Are you ready? Remember, look at the pictures first. (The preceding sentence

TABLE 1

MEAN NUMBER OF ERRORS TO A CRITERION OF 10 CONSECUTIVE CORRECT RESPONSES WITH GENERAL INSTRUCTIONS

Intelligence	Groups						Means by IQ
	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Average IQ	71.2	21.0	38.9	34.5	14.3	22.2	41.5
High IQ	57.1	27.8	19.7	17.0	6.5	11.4	27.8
Means by Age	64.2		29.3		10.4		

Note.—$F_{Age} = 37.66$, $df = 2$, $P < .001$; $F_{Intel.} = 7.14$, $df = 1$, $P < .01$; $F_{Age \times Intel.} = <1$.

was repeated together with the following sentence whenever necessary.) Remember, there is an idea that will tell you which picture will give you the marble.

The procedure was continued until *S* made 10 consecutively correct responses, which was the criterion of success. At this point *E* commented on *S*'s good work and asked him how he knew where the marble was each time. If *S* could not state the rule correctly the procedure was continued with questions following each block of 10 consecutively correct responses until *S* gave a correct answer or completed 150 trials. Any *S* who did not achieve criterion performance at that point was considered to have failed. A second session, held within 2 weeks of the first, was devoted to the administration of the WISC.

Experimental design.—In both experiments a 3 × 2 factorial design was used with three ages and two intelligence levels constituting the respective variables. Thus there were six experimental groups with 15 *Ss* in each group, a total of 90 *Ss* in each experiment.

RESULTS

The results were analyzed in terms of three measures of performance: (*a*) the mean number of errors to criterion, (*b*) the number of *Ss* who attained the criterion, and (*c*) the number of *Ss* who verbalized the concept at the point of success.

Experiment I.—The mean number of errors per group and the results of an analysis of variance of errors,[2] presented in Table 1, indicate that age and intelligence were associated with reliable differences in performance, thus reproducing the findings reported by Osler and Fivel (1961). Similar results were obtained from an analysis of the data on frequency of *Ss* who attained criterion performance and those who were able to verbalize the concept, as shown in Table 2. In summary, then, it turned out that with instructions that did not specify the nature of the task, age and intelligence were associated with improved achievement in concept attainment on each of three measures of performance.

Experiment II.—Table 3 presents the mean number of errors per group and the results of an analysis of variance of errors, respectively, for Exp. II. In this experiment, in which the nature of the task was explicitly defined, the effect of intelligence disappeared; only age was associated

[2] Among the 14-yr-old *Ss* the conditions necessary for application of an analysis of variance were not entirely satisfied because of skewness of the distribution and the smaller variance of the 14-yr-old High Intelligence group. For this age group the distribution of errors was examined and compared (by χ^2) for average and high intelligence within and between the two experiments. There was no evidence of significant differences among the 14-yr-old *Ss* according to intelligence or between the experiments.

TABLE 2

NUMBER OF *Ss* WHO REACHED CRITERION AND NUMBER WHO VERBALIZED THE CONCEPT UNDER GENERAL INSTRUCTIONS

Intelligence	Groups						All Ages	
	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*			
	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.
Average IQ	4	1	8	2	13	13	25	16
High IQ	7	4	14	11	15	14	36	29
Total	11	5	22	13	28	27	61	45

Note.—Criterion data: Age $\chi^2 = 22.690$, $df = 2$, $P < .001$; Intelligence $\chi^2 = 6.156$, $df = 1$, $P < .02$. Verbalization data: Age $\chi^2 = 33.068$, $df = 2$, $P < .001$; Intelligence $\chi^2 = 7.512$, $df = 1$, $P < .01$.

with significant differences in the number of errors. Similar results were obtained from an analysis of the other two measures of performance, number of *Ss* who achieved criterion performance and number of *Ss* who verbalized the concept, as shown in Table 4.

Combined data from the two experiments.—In view of the comparable variances obtained in both experiments the error data were combined and analyzed for a possible Instructions × *S* interaction. The results of this analysis, presented in Table 5, show that while instructions produced no main effect on performance, the Age × Instruction interaction was significant at the .05 level. When the data were subsequently analyzed separately for each age group, it turned out that only for the 6-yr. *Ss* did instructions produce a significant difference in performance ($F = 5.92$, $df = 1.59$, $P < .05$) in favor of explicitness. Another finding, rather surprising, and one difficult to explain, was that the bright 10-yr. *Ss* were *handicapped* by the explicit instructions. The mean number of errors for these *Ss* under the two conditions was 19.7 and 37.0, respectively ($t = 2.57$, $df = 28$, $P < .02$). The oldest group, the 14-yr. *Ss*, performed identically under both conditions. However, as is apparent from their scores, these *Ss* performed near the maximum possible level, and, therefore, could not

TABLE 3

MEAN NUMBER OF ERRORS PER GROUP FOR REACHING CRITERION OF 10 CONSECUTIVE CORRECT RESPONSES WITH EXPLICIT INSTRUCTIONS

Intelligence	Groups						Means by IQ
	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Average IQ	46.1	25.5	34.2	26.5	14.1	17.4	31.5
High IQ	46.5	34.0	37.0	31.3	7.7	17.4	30.4
Means by Age	46.3		35.6		10.9		

Note.—$F_{Age} = 13.52$, $df = 2$, $P < .001$; $F_{Intel.} < 1$; $F_{Age \times Intel.} < 1$.

TABLE 4

NUMBER OF *Ss* WHO REACHED CRITERION AND NUMBER WHO VERBALIZED THE CONCEPT UNDER EXPLICIT INSTRUCTIONS

Intelligence	Groups						All Ages	
	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*			
	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.
Average IQ	11	6	12	8	14	12	37	26
High IQ	9	4	9	9	14	14	32	27
Total	20	10	21	17	28	26	69	53

Note.—Criterion data: Age $\chi^2 = 7.081$, $df = 2$, $P < .05$; Intel. $\chi^2 = 1.552$, $df = 1$, *ns*. Verbalization data: Age $\chi^2 = 17.715$, $df = 2$, $P < .001$; Intelligence $\chi^2 = .046$, $df = 1$, *ns*.

be expected to show improvement under any conditions.

The analysis of the combined frequency data of successful *Ss* shows that aside from age, reliable differences were produced by instructions in the Average Intelligence but not in the High Intelligence group. The same results were yielded by the data on correct verbalization.

Rank-order correlations were computed between error scores and MAs in each experiment. The rho values obtained were −.68 under general instructions and −.51 under explicit instructions. Both values are significant at the .001 level. Since the MA is heavily determined by the CA, the value of the correlation per se does not throw light upon the role of the IQ alone. However, as *Ss* in both experiments were identical in age and intelligence, the difference between the two values of rho may be attributed to variation in experimental conditions. Unfortunately, no statistical test exists for evaluating the difference between the rho values; therefore all that can be said is that the direction of the difference is consistent with the other findings.

TABLE 5

ANALYSIS OF VARIANCE OF ERRORS FOR EXP. I AND II COMBINED

Source	*df*	*MS*	*F*
Instructions (I)	1	616.05	
Age (A)	2	29,819.62	45.05**
Intelligence (IQ)	1	2,456.81	3.71
I × A	2	2,382.52	3.60*
I × IQ	1	1,798.67	2.72
A × IQ	2	7.90	
I × A × IQ	2	407.87	
Within	168	661.89	
Total	179		

* $P < .05$.
** $P < .001$.

DISCUSSION

Under general instructions high intelligence was associated with superior achievement in all three measures of performance; but this superiority was not maintained under explicit instructions. If the assumption that specific instructions eliminate the problem-finding component from the task is correct, then the conclusion to be drawn is that high intelligence gave *S* an advantage in problem definition but not in problem solution in the concept attainment task. It is entirely possible, however, that were the concept task one of greater difficulty, intelligence might also prove to be advantageous in the solution phase of the task. This requires further investigation.

Since age was associated with more effective concept attainment on all three

measures of performance and under both types of instructions, it may be concluded that maturation involves growth in problem *solving* ability at least until some point between 10 and 14 yr.

In order to assess changes in problem *finding* ability as a function of age, it is necessary to compare performance at each age level under the two types of instructions. That 6-yr. *Ss* improved under the explicit instructions is apparent from the significant Age × Instruction interaction in Table 5. The situation was not clear for the 10-yr. *Ss* because they actually deteriorated under explicit instructions. As this finding seemed to be inexplicable on any logical basis and was inconsistent with the results pertaining to the other two age groups, it was decided to replicate this portion of Exp. II. Upon replication it turned out that the average 10-yr. *Ss* profited from the explicit instructions, while the superior *Ss* were unaffected. As to the 14-yr. *Ss*, whose performance was the same under both conditions, they did not have much room for improvement, precluding the possibility of reaching a definite conclusion on the effect of instructions at this age.

It thus appears that at least between 6 and 10 yr. *Ss* benefited from explicitness of instructions. It is therefore reasonable to conclude that the more effective performance associated with increasing age is not brought about by improvement in problem finding, but rather by more effective problem solving. This points to a possible interesting difference between mental age and intelligence, the effectiveness of the latter being due to more effective problem finding rather than problem solving.

Summary

A concept attainment task was presented under general and explicit instructions to *Ss* 6, 10, and 14 yr. of age, divided into two intelligence groups. The results were analyzed in terms of three measures: errors, number of *Ss* who achieved the criterion of success, and number of *Ss* who verbalized the concept correctly.

Under nonspecific instructions superior intelligence was associated with more effective concept attainment on three measures of performance, but under explicit instructions *Ss* of average intelligence improved while those of superior intelligence remained unchanged. Apparently, superior intelligence gave *Ss* an advantage in the problem finding phase of the task, but not in actual problem solution.

Performance improved throughout the age range studied. There was no evidence for improved problem finding ability between ages 6 and 10 yr., and it was difficult to evaluate the data for the 14-yr-olds. It appears, therefore, that problem solving ability, but probably not problem finding ability, increases within the age range studied.

References

Osler, S. F., & Fivel, M. W. Concept attainment: I. The role of age and intelligence in concept attainment by induction. *J. exp. Psychol.*, 1961, **62**, 1–8.

Osler, S. F., & Powell, M. G. Apparatus for the study of discrimination and concept formation. *Amer. J. Psychol.*, 1960, **73**, 627–629.

(Early publication received December 18, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 534–537

INHIBITORY POTENTIAL IN ROTARY PURSUIT ACQUISITION BY NORMAL AND DEFECTIVE SUBJECTS[1]

R. WAYNE JONES AND NORMAN R. ELLIS

Louisiana State University *George Peabody College for Teachers*

One systematic approach to the problem of behavioral subnormality is represented by attempts to utilize descriptive or "explanatory" models used in the study of the normal organism. Constructs such as habit strength, motivation, inhibition, generalization, etc., may serve to delimit a class of functions within which individual differences may be most evident. For example, it is possible that the defective's handicap involves mainly an associative mechanism. On the other hand, the principal deficit may be motivational with associative and other components contributing secondarily if at all. The present experiment adopts this general orientation. The concern here is with inhibition.

A previous study (Ellis, Pryer, & Barnett, 1960) showed that normal *Ss* gained relatively more during a rest interval than did mental defectives while performing on the rotary pursuit task. The defining operations for this gain were similar to those used by other theorists (Ammons, 1947; Hull, 1952; Kimble, 1949) to describe the dissipation of inhibition. On this basis, it was hypothesized, ad hoc, that the defective built up less inhibition (I_R) than did the normal *S*. Hull's model was used since the general notion of fatigue seemed most germane. From this hypothesis Ellis et al. (1960) predicted that spaced as compared to massed learning should be *relatively* more favorable to normal than to defective *Ss*. The present experiment is designed to test the hypothesis that there are differences in the rate of accumulation of inhibition in these *Ss* which are independent of performance skill. These differences should appear as interaction effects.

METHOD

Subjects.—The 160 *Ss* consisted of 80 institutionalized mental defectives and 80 normal high school students. Equal numbers of males and females occurred in each group. The defectives ranged in IQ (Binet or Otis) from 60 to 75, and in CA from 13 to 28 yr. Those relatively free of gross motor and sensory defects were selected. The IQs of the tenth grade high school students were not available, but only those with a grade average above "C" were used. They ranged in CA from 14 to 16 yr.

Apparatus.—The apparatus was a pursuit rotor (Lafayette Instrument Company, Model No. 203) and an electric stopclock calibrated in .01 sec. The rotor turntable was 10 in. in diameter with a ¾ in. round brass target mounted flush in the surface 3½ in. from the center. The turntable rotated clockwise in a horizontal plane at 60 rpm. The on-off periods were controlled automatically by a repeat cycle timer.

Procedure.—The 80 *Ss* in each group were assigned equally and without bias to two treatments. Sexes were represented equally in each of the four subgroups. Massed practice was defined by 10-sec. intertrial intervals

[1] This study was supported by Grants M-2229 and M-4433 from the National Institute of Mental Health, Public Health Service, made to the second author. A thesis based on this investigation was submitted by the first author to the graduate faculty, Louisiana State University, in partial fulfillment of the Masters of Arts degree. Thanks are due Margaret W. Pryer, Shirley Wrightsman, and Janice Robinson for statistical computations.

and distributed practice by 30-sec. intervals. A trial was 20 sec. practice. Each *S* received a block of 30 trials, a 5-min. rest, and a postrest block of 20 trials. The *Ss*' behavior was not controlled during the rest except, of course, they were not permitted to engage in any observable type of practice. General encouragement was given throughout practice.

In order to avoid "floor" effects in performance, *Ss* failing to score as much as 3 sec. on target for one of the last two trials in the first block were excluded. This criterion eliminated 71 defectives and 4 normal *Ss*. The defectives were 16 males and 39 females in the massed group and 5 males and 11 females in the spaced group. The 4 normal *Ss* were females and in the massed group.

Results and Discussion

Figure 1 depicts the performance of the four groups for the pre- and postrest trials. All curves are fairly typical for this task, consisting of linear prerest improvement, a reminiscence effect, and postrest warm-up. Table 1 presents a summary of an analysis of variance of Prerest Trials 11–30.

The A × B interaction shown in Table 1 is critical to the hypothesis of a differential build-up of I_R. The F of 3.67 approaches the .05 level (3.92 required with $df = 1/150$). Other results summarized statistically in Table 1 are predictable from previous work in this area. The sex effect will receive attention later.

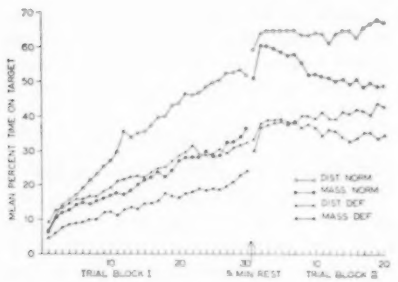

Fig. 1. Prerest and postrest performance curves.

TABLE 1

Analysis of Variance of Prerest Trials 11–30

Source	*df*	*MS*	*F*
Normals vs. Defectives (A)	1	49,124	40.84**
Massed vs. Distributed (B)	1	60,077	49.95**
Male vs. Female (C)	1	32,382	26.92**
A × B	1	4,416	3.67
A × C	1	3,029	2.52
B × C	1	6,843	5.69*
A × B × C	1	44	0.04
Error	152	1,202	
Between *Ss*	159		
Trials (D)	19	1,988	108.60**
A × D	19	73	4.00**
B × D	19	86	1.81*
C × D	19	33	0.59
A × B × D	19	10	0.70
A × C × D	19	12	0.70
B × C × D	19	12	1.14
A × B × C × D	19	20	1.11
Error	2888	18	
Within *Ss*	3040		
Total	3199		

* $P < .05$.
** $P < .01$.

Table 2 presents reminiscence scores. Table 3 summarizes an analysis of variance of reminiscence scores where reminiscence is defined as the difference between the last prerest and the third postrest score. Again, the A × B interaction is central to the hypothesis. Its absence here indi-

TABLE 2

Gain in Percentage of Time on Target (Reminiscence) after 5-Min. Rest

Subjects	*N*	Massed		Distributed	
		Mean	*SD*	Mean	*SD*
Normals					
Men	40	4.47	2.01	2.14	2.33
Women	40	5.08	1.65	2.94	3.44
Defectives					
Men	40	2.27	2.51	1.62	2.40
Women	40	3.01	3.37	0.69	2.35

TABLE 3

ANALYSIS OF VARIANCE OF REMINISCENCE

Source	*df*	*MS*	*F*
Normals vs. Defectives (A)	1	123.45	23.03**
Massed vs. Distributed (B)	1	138.05	25.75**
Male vs. Female (C)	1	3.75	0.69
A × B	1	5.75	1.07
A × C	1	26.80	5.00*
B × C	1	5.40	1.01
A × B × C	1	26.24	4.89*
(Error *MS*)	152	5.36	

* $P < .05$.
** $P < .01$.

cates that reminiscence in normals and defectives is similarly affected by degree of distribution of practice.

Table 4 presents a summary of an analysis of variance of the 20 postrest trials. In this instance, the A × B interaction is highly significant (.01 level) which is generated by a larger difference between the normals' curves compared to those for the defectives. Possibly, this effect could be attributed to conditioned inhibition (sI_R). The greater associative capacity of the normals as well as the development of more I_R makes this interpretation tenable. The decay segment slope of the massed-practice curve for the normals is slightly steeper than that for the defectives. This trend is in accord with the prediction since it suggests a greater build-up of inhibition in the normal group. However this finding is not supported statistically as shown by the absence of an A × B × D interaction in Table 4.

TABLE 4

ANALYSIS OF VARIANCE OF POSTREST TRIALS 1–20

Source	*df*	*MS*	*F*
Normals vs. Defectives (A)	1	30,708	1668.12**
Massed vs. Distributed (B)	1	21,923	279.79**
Males vs. Females (C)	1	11,534	147.20**
A × B	1	2,891	36.90**
A × C	1	1,300	16.60**
B × C	1	5,493	70.11**
A × B × C	1	1,981	25.29**
Error *MS*	152	78	
Between *Ss*	159		
Trials (D)	19	186	2.03**
A × D	19	69	0.76
B × D	19	240	2.61**
C × D	19	20	0.22
A × B × D	19	31	0.34
A × C × D	19	27	0.30
B × C × D	19	22	0.24
A × B × C × D	19	19	0.22
Error *MS*	2888	92	
Within *Ss*	3040		
Total	3199		

** $P < .01$.

Marked sex differences are evident in these analyses. In prerest practice the females within each of the four subgroups show poorer performance than the males. There are larger differences between the sexes under distributed practice (B × C interaction) and the males show greater gain with trials (C × D) independent of degree of distribution. The females show slightly larger amounts of reminiscence though this effect is not significant. (See main effect in Table 3.) The A × C interaction derives from the female normals showing more reminiscence than the normal males and a reversal of this trend within the defective group. Also, there is a reversal of this effect due to distribution of practice in the male and female defectives as compared to the trend for the male and female normals (A × B × C).

In postrest practice the females' performance, in general, falls far below that of the males with the exception of the massed defectives. In this case the performance of these *Ss* is consistently above that of the defective males. The massed normal males are consistently above the females.

This reversal is mainly responsible for the A × B × C effect. The differences between massed and distributed practice were larger in the males than in the females, thus the B × C interaction. Likewise the differences between the sexes were greater in the normals than in the defectives. Probably the sex effects are attributable to a combination of motivational and motor skill factors.

Over-all these data suggest that normal *Ss* build up more inhibition than do defectives although this finding is not independent of skill. This is supported by the near significant A × B interaction in prerest practice, larger amounts of reminiscence occurring in the normals, the A × B interaction in postrest performance, and the higher rate of decay trend in the massed normal group as compared to the defectives in postrest performance. This generalization requires considerable qualification, however. In pre- and postrest practice the A × B and A × B × D interactions are the only indices to differential build-up of inhibition. Such a prediction hinges upon the assumption that distribution of practice influences inhibition rather than associative mechanisms. Thus if inhibition accumulates at different rates in the normal and defective, these interaction effects could be predicted. The prerest A × B effect is near statistical significance and suggests that I_R may accumulate differentially in these *Ss*, independently of performance level. The postrest effect is statistically significant but in this instance, a factor other than "temporary" inhibition seems to be operative. In the reminiscence analyses an A × B interaction is again predictable. It did not occur and the conclusion that these groups are not differentially influenced by distribution of practice is necessary. An actual decay in performance such as that seen in the massed groups in postrest practice could have provided fairly clear evidence for a differential inhibition build-up but the trend proved insignificant statistically. However, the factor of "selective sampling" lends some credibility to the hypothesis of a differential build-up of I_R. By eliminating *Ss* with low performance the defective and normal groups were made more similar on the criterion measures.

On the basis of these data it seems likely that normal *Ss* accumulate more inhibition than do defective *Ss*. This conclusion is supported principally by the A main effect of the reminiscence analysis. However, this effect seems proportional to performance level and degree of distribution of practice. An hypothesis of a differential build-up of inhibition in these *Ss* independent of performance level is not presently tenable. Further experimentation with more homogeneous groups of defective and normal *Ss* should be carried out in view of trends which favor the hypothesis.

Summary

The present experiment compared normals and defectives under conditions of massed and distributed practice on the rotary pursuit task. Over-all, normal *Ss* showed a larger build-up of inhibition than the defectives although this effect could not be demonstrated independently of performance level and distribution of practice. There were sex differences in both prerest and postrest performance. There were some characteristics of postrest performance which could be ascribed to the development of a "permanent" inhibition effect.

References

Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *J. exp. Psychol.*, 1947, **54**, 263–281.

Ellis, N. R., Pryer, M. W., & Barnett, C. D. Motor learning and retention in normals and defectives. *Percept. mot. Skills*, 1960, **10**, 83–91.

Hull, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

Kimble, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.

(Received January 21, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 538–545

THE GENERALITY OF THE "CONDITIONING AXIOM" IN HUMAN BINARY PREDICTION [1]

PETER L. DERKS [2]

University of Pennsylvania

The linear learning models of Estes (1950) and Bush and Mosteller (1951a, 1951b, 1955) have been successful in describing the results of most binary prediction experiments. When the reward and nonreward parameters of a stochastic model are equal, the probability of the choice of an event approaches the probability of its occurrence. This "probability matching" has been reported by several investigators (Grant, Hake, & Hornseth, 1951; Hake & Hyman, 1953; Humphreys, 1939; Jarvik, 1951). However, if the nonreward parameter has no effect, the probability of the choice of the more frequent event (A) approaches one. Subjects have been found to predict A on nearly every trial (Goodnow, 1955a, 1955b; Goodnow & Postman, 1955). Thus either result is described by linear learning models with proper restrictions on their parameters.

One of the basic axioms of these models is the "conditioning axiom." In the context of the present experiments, this axiom asserts that the occurrence of an event will increase the probability that *S* will predict that event to occur again. As the event is repeated, the probability that it will be predicted increases monotonically.

To test this assertion, Estes (1957) asked *S*s to predict two nonsense syllables presented in 66:34 proportion for 2000 trials. For the last 1500 trials, the proportion of predictions of A increased monotonically as a function of three successive presentations of the syllable. However, several experiments have shown that *S*s avoid long runs in the prediction of random events (Goodfellow, 1938; Goodnow, 1955b; Skinner, 1942; Smith, 1949). Nicks (1959) reported that for event schedules of 75:25 the proportion of A predictions increased only until A had occurred twice in succession and then decreased. Anderson and Whalen (1960) obtained similar results with other schedules. These *S*s did not respond in accordance with the prediction based upon the conditioning axiom but behaved as if they believed the "gambler's fallacy" described by Jarvik (1951).

Thus, although either probability matching or 100% choice of A can be predicted from linear learning models, the changes in proportion of choice with successive occurrences of an event are not always in agreement with the monotonic increase expected from the conditioning axiom basic to the models.

EXPERIMENT I

Galanter and Gerstenhaber (1956) suggested that *S* attempts to reduce the "cost" of each prediction. If knowledge of the correct alternative is contingent on *S*'s prediction of that alternative the proportion of A choices is greater than if the correct alternative is indicated on every trial (Detambel, 1955; Goodnow, 1955a; Neimark, 1956). When *S* receives money

[1] This research was supported in part by the National Institute of Mental Health Grant M-2293 and the Psychology Department of the University of Pennsylvania.

[2] Now at the College of William and Mary.

for correct prediction the A choice proportion approaches 100% (Edwards, 1956; Galanter & Smith, 1958; Siegel & Goldstein, 1959).

Although temporal variables are important in most learning experiments, manipulation of intertrial interval has little effect on choice proportion (Grant, Hornseth, & Hake, 1950). Outcome delay and response latency have not been examined as variables in human binary prediction. The following experiment investigated the influence of manipulations of these parameters on the proportion of choice as a function of the successive occurrence of an event and compared the results with the function asserted by the conditioning axiom.

Method

Apparatus.[2]—The equipment used in this experiment was a modification of that described by Bush and Mosteller (1955, p. 277). The *S* was seated at the panel illustrated in Fig. 1. In this study he was concerned only with the following aspects of the panel. An amber stimulus light (labeled S) on the upper panel informed *S* of the beginning of each trial. This light remained on for the duration of the trial and flashed off for 0.5 sec. indicating the end of a trial. Two white lights (labeled R and L) on the lower panel were the events. The *S* pressed one of two response buttons (labeled R′ and L′) below the event lights to predict which event would occur on a given trial. The light did not appear until after *S* had made a prediction. A latency counter (labeled LC) could register the time in 0.1 sec. between the onset of the stimulus light and the depression of the response button. This counter was operated for only one of the experimental groups.

The *E* was in a separate room where the display viewed by *S* was reproduced. All elements of the panel were the same except the response buttons which were represented by two lights. A red light on the right of *E*'s panel indicated the right response button and a blue light on the left indicated the left response button. When *S* pressed the right or left button, the corresponding light came on in *E*'s room and *E* recorded the prediction.

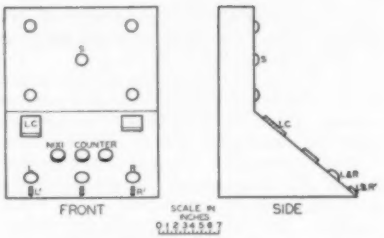

FIG. 1. Schematic diagram of human choice apparatus with S—stimulus light, L—left event light, R—right event light, L′—left response button, R′—right response button, LC—latency counter, and the nixi counter.

Time intervals were controlled by two mechanical timers. The "outcome delay" timer controlled the interval between *S*'s response and the onset of the event light. The "intertrial interval" timer controlled the duration of the event light and the subsequent onset of the stimulus light. A latency counter registered 0.1 sec. between the onset of the stimulus light and *S*'s response. Unless stated otherwise the intertrial interval and outcome delay were 1 sec. Accurate latency records were not kept as the record of *S*'s predictions required most of *E*'s attention.

The schedules of stimulus and event lights were programed on punched tapes which were stepped through two Western Union 23-A tape transmitters. A pulse from the outcome delay timer stepped the transmitter that carried the event schedule. The transmitter that carried the stimulus schedule was stepped by the intertrial interval timer or externally by *E* to position the tape. The same schedule of events was used for all *S*s. It was selected from the Rand (1955) table of a million random digits by the designation of 00 to 74 as A, the right light, and 75 to 99 as B, the left light. The series was 250 events in length and contained 187 rights and 63 lefts. There were 11 A runs, 6 trials or more in length.

Subjects.—Eighty male undergraduates at the University of Pennsylvania served as *S*s as a part of their course requirements in elementary psychology and for the monetary reward described below. They were assigned to eight groups on the basis of their order of appearance to be tested. The first 10 *S*s were Group 1, the second 10 *S*s, Group 2, etc.

Procedure.—Each *S* for all groups was instructed to guess which of two lights would

[2] The apparatus was designed and constructed under the direction of E. H. Galanter by Biophysical Instruments, Incorporated, of Germantown, Pennsylvania.

come on following the appearance of the stimulus light. The guess was made by pushing the button under the light.

The *E* gave *S* an overestimate of the length of the experiment both in time and number of trials. Questions pertaining to the instructions were answered and *E* left the room.

Groups.—Group 1 (Control) was a replication of the experiments mentioned earlier on human binary prediction of a 75:25 schedule. In Group 2 (Outcome Delay) the outcome delay was 15 sec.; in Group 3 (Intertrial Interval) the intertrial interval was 15 sec.

In Group 4 (Restricted Latency) an attempt was made to limit the response latency. The general instructions were given with the addition that each guess was to be made "as quickly as possible" after the stimulus light flashed on. The latency counter in *S*'s room operated for this group. Since this counter, as it registered, clicked every 0.1 sec., *S* could respond to an auditory cue as well as the stimulus light. The *S* was told that he should respond in about 0.2 sec. The *S*s who required more than 0.5 sec. on over 25% of the trials were not included in the group. The average latency for the other groups was approximately 5 sec. Thus any *S* who responded at the required rate spent relatively little time on his decision. Nine *S*s failed to meet the criterion before the group of 10 was completed.

In Group 5 (Reward), *S* gained 5 cents for every correct response. The nixi counter registered once for each nickel and displayed the total amount won. The *S* was instructed concerning this tabulation with the additional reminder that "1 equals a nickel, 2 equals a dime, 5 equals a quarter, 20 equals a dollar, etc." In 250 trials these *S*s earned from \$7.05 to \$8.75.

In Group 6 (Risk), *S* gained 5 cents for every correct response and lost 5 cents each time he was wrong. The nixi counter displayed the net total. The *S* was instructed concerning this tabulation plus the same reminder given to Group 5. These *S*s earned from \$3.25 to \$5.50 in 250 trials.

In Group 7 (Contingent) the same 75:25 schedule was used, but the event light only appeared if it was predicted correctly. The *S* was instructed:

> Try to push the correct button as often as possible. These buttons are independent of each other. There are trials on which both buttons are correct, trials on which neither button is correct, and trials on which only one button is correct. Nevertheless, you may push only one button per trial. When a correct button is pushed the light above it will come on. Even if both buttons are correct only the light over the button you push will come on. Thus you only receive information about one button at a time.

In Group 8 (Contingent-Risk) the 75:25 schedule was again used with the contingent procedure described for Group 7, but *S* received 5 cents when correct and lost 5 cents when wrong with the total presented on the nixi counter. These *S*s earned from \$3.20 to \$5.70.

Results

Probability matching.—Differences in probability matching are usually determined from the average of the final choice proportions of a group of *S*s. However it would be possible for 1 *S* to predict A half the time on a 75:25 schedule while another predicted that event on every trial. The average data for the 2 *S*s would be a probability match while the individual results would not. To overcome this source of error each group was also described in terms of individual *S*s who were probability matchers (PMs).

A PM was defined as any *S* whose proportion of A predictions on the last 100 trials was within ± 2 *SD* of the actual proportion of that event. The *SD* estimated by:

$$\hat{\sigma} = \sqrt{N(\pi_1 \cdot \pi_2)} \qquad [1]$$

where $\hat{\sigma}$ = estimated *SD*, π_1 and π_2 = the proportion of A and B, and N = number of trials.

For a 75:25 schedule 2 *SD* = approximately 9. Thus any *S* who predicted A from 66 to 84 times on the last 100 trials was considered to be a PM. The number of PMs and the mean number of A predictions on the last 100 trials for each group is summarized in Table 1.

Extension of the intertrial interval or the outcome delay to 15 sec. is not sufficient to significantly alter choice

proportion or number of PMs. Group 1 (Control), Group 2 (Outcome Delay), and Group 3 (Intertrial Interval) are not significantly different. These results agree with other binary prediction experiments in which intertrial interval was a variable.

However Group 4 (Restricted Latency), when compared with Group 1, has significantly more PMs ($P < .05$, $B = 9$, $D = 4$) from the Finney table of the Fisher test of χ^2 (Siegal, 1956) and a lower proportion of A predictions on the last 100 trials ($P < .05$, $t = 2.45$, $df = 18$). Furthermore, of the 9 *Ss* who were not included in this group because they failed to respond within 0.5 sec. on 25% of the trials, only 3 were PMs. Therefore, speed instructions resulted in an increase in PMs if *S* responded to the instructions by making decisions at the average rate of approximately .5 sec. per decision (the criterion for inclusion in Group 4), but did not result in an increase of PMs in the case of *Ss* who responded to the speed instructions by speeding up their decisions (as compared with Group 1, which worked at the rate of approximately 5 sec. per decision), but only to an average rate between .5 and 3.0 sec. per decision.

When incentive is manipulated, Group 6 (Risk) has the highest proportion of A predictions, followed in order by Group 5 (Reward) and Group 1. These results are in the direction reported in other studies but the differences here are not significant. Although *Ss* in Groups 5 and 6 all reported that they believed they were going to be paid they did not significantly increase their choice proportion over *Ss* who received no money. Numbers that registered on a counter did not affect behavior as did actual nickels or chips for money of other studies (Bush & Mosteller, 1955, Ch. 13; Edwards, 1956; Galanter & Smith, 1958; Siegel & Goldstein, 1959). The reason for this difference is not clear.

TABLE 1

PROBABILITY MATCHERS (PMs) AND PROPORTION OF MORE FREQUENT EVENT PREDICTIONS ($\bar{P}_1$) ON FINAL 100 TRIALS

Group	Condition	No. of PMs	$\bar{P}_1$
1	Control	4	.850
2	Outcome Delay	5	.820
3	Intertrial Interval	5	.885
4	Restricted Latency	9	.764
5	Reward	3	.879
6	Risk	3	.880
7	Contingent	5	.803
8	Contingent-Risk	1	.948

The *Ss* in the contingent groups (Groups 7 and 8), when questioned after the experiment, all said they accepted the independence of the events. Verbal estimates of "proportion correct" for each event did not sum to 1 for 13 of the 20 *Ss*. Since none of these groups had more than 5 PMs out of 10 *Ss*, the differences between number of probability matchers are not significant. However, Group 8 (Contingent-Risk) had a significantly greater proportion of A predictions than Group 1 ($P < .05$, $t = 2.48$, $df = 18$) and Group 7 (Contingent) ($P < .01$, $t = 3.86$, $df = 18$).

These results agree with studies which reported that a contingent-risk procedure resulted in a greater A choice proportion than other procedures. However, in this study, the contingent procedure did not result in a significant difference in choice proportion from a noncontingent one when *Ss* were unrewarded.

Choice and event repetition.—The proportion of choice as a function of successive occurrence of an event was calculated from *Ss*' responses to sequences of six or more repetitions of

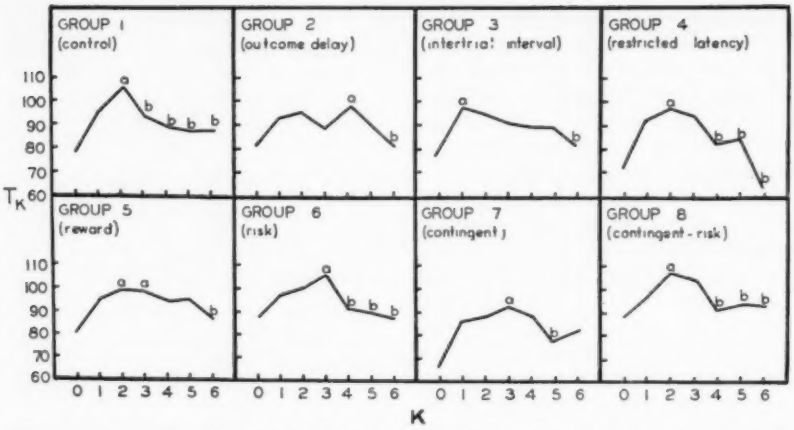

FIG. 2. T_k as a function of k in Exp. I: a—Maximum T_k; b—Subsequent $T_k > 2$ *SD* below maximum.

A.[4] There were 11 of these runs in the 250-trial schedule. The predictions of A after it had occurred 0, 1, 2, . . ., 6 times in a row were summated over the 10 *S*s in each group for the 11 sets of six successive A events. Thus the sample consisted of 110 cases in which A followed B either 0, 1, 2, . . ., 6 times in succession in the sets of A runs that were six or longer. This is:

$$T_k = \sum_{i=1}^{10} \sum_{j=1}^{11} n_{kji} \qquad [2]$$

where:

i = any individual subject ($i = 1, 2, \ldots, 10$)

j = any set of six successive occurrences of A ($j = 1, 2, \ldots, 11$)

k = any number of successive occurrences of A within a succession of six or more such occurrences ($k = 0, 1, \ldots, 6$)

n_{kji} = number of A predictions following k successive occurrences of A for set j and subject i ($n = 0, 1$)

T_k = total number of A predictions following k successive occurrences of A for all 11 sets and all 10 *S*s (T_k is a positive integer, $0 \leq T_k \leq 110$)

Figure 2 summarizes T_k as a function of k for all groups. The conditioning axiom states that T_k will increase monotonically as k increases, i.e., the predictions of A will increase with its successive occurrence. If there is a T_k more than 2 *SD* less than any preceding T_k the function cannot be described as monotonically increasing. Equation 1 gives an estimate of 2 *SD* for T_k of approximately 9. In Fig. 2 the maximum T_k and any subsequent T_k that is 2*SD* less than the maximum is noted. T_k is not a monotonically increasing function of k for any group.

EXPERIMENT II

Edwards (1961) found that for schedules of 1000 trials duration "the asymptotic probability of prediction was more extreme than the probability of occurrence of the event pre-

[4] This statistic was suggested by R. R. Bush and H. C. Morlock of the University of Pennsylvania in an unpublished memorandum.

dicted." He computed the proportion of choice of an event as a function of its successive occurrence using the method described by Nicks (1959). On the first 400 trials the function rises to a peak value after which the proportion of predictions of A decreases. However, on the last 600 trials the proportion of choice of an event increases monotonically as a function of the successive occurrence of that event as asserted by the conditioning axiom. Lindman and Edwards (1961) have found that this change is already underway after 100 trials with a 50:50 schedule. As mentioned earlier Estes (1957) found that the function increased monotonically on Trials 500 to 2000.

Experiment II investigated the proportion of choice of A as a function of its successive occurrence in a 1000-trial sequence. In order to examine any progressive change in this function T_k was computed for successive blocks of 250 trials.

Method

The same equipment described in Exp. I was used. Ten male undergraduates (Group 9) at the University of Pennsylvania were paid $4.00 each to serve as *S*s. They were given the same instructions cited in Exp. I. The length of the experiment was given as "1000 trials and about 2 hr." Each *S* predicted the same 75:25 noncontingent schedule described in Exp. I. The 250 trials of this schedule were repeated four times without interruption to make 1000 trials.

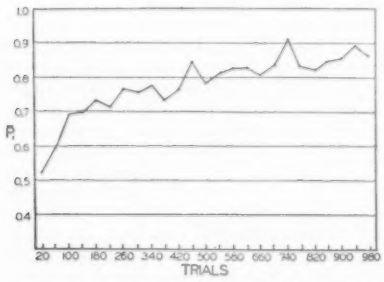

FIG. 3. Proportion of predictions of the more frequent event on a 75:25 schedule ($\bar{P}_1$) as a function of trials in blocks of 40.

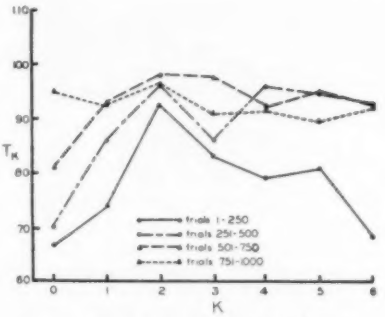

FIG. 4. T_k as a function of k in Exp. II.

Results

The proportion of predictions of A as a function of trials in blocks of 40 is presented in Fig. 3. On this 75:25 contingent schedule the choice proportion continued to increase throughout 1000 trials. However, 3 *S*s were still PMs on the last 100 trials. Although group choice proportion does not stabilize in 1000 trials, individual *S*s may consistently probability match.

T_k as a function of k is summarized for each quarter of the sequence in Fig. 4. On Trials 1–250, T_k is the "peaked" function of k found in the earlier experiments of this study. A curve that increases monotonically can fit T_k as a function of k for all but one point for Trials 251–500 (T_k for $k = 3$ is more than 2 *SD* below the peak value at $k = 2$) and all points for Trials 501–750. However, T_k is a linear function of k for Trials 751–1000 and the slope is approximately zero.

In order to examine the cause of the change in T_k, some of the information reflected by T_k was tabulated more fully. The mean run length of predictions of A and the number of predictions of B following its occurrence are summarized in Table 2. The run length of A pre-

TABLE 2

MEAN RUN LENGTH OF PREDICTIONS OF EVENT A ($\bar{R}_1$) AND TOTAL NUMBER OF PREDICTIONS OF EVENT B FOLLOWING ITS OCCURRENCE (r_2) FOR FOUR SUCCESSIVE 250-TRIAL BLOCKS

Prediction	Trials			
	1–250	251–500	501–750	751–1000
$\bar{R}_1$[a]	3.93	5.70	10.20	17.51
r_2[b]	228	184	155	91

Note.—Event B occurred 630 times for the 10 *Ss* in each 250-trial block. All t values were obtained by the Lord technique as described by Mosteller and Bush (1954).

[a] For $\bar{R}_1$, Trials 751–1000 differ from 1–250 ($P < .02$, $t = 3.15$, $df = 9$) and 251–500 ($P < .05$, $t = 2.64$, $df = 9$), and Trials 1–250 differ from 501–750 ($P < .02$, $t = 2.75$, $df = 9$). None of the other differences are significant.

[b] For r_2, Trials 751–1000 differ from 1–250 ($P < .01$, $t = 3.88$, $df = 9$) and 251–500 ($P < .02$, $t = 2.80$, $df = 9$). None of the other differences are significant.

dictions increases from quarter to quarter while the B predictions that follow A's occurrence decrease. The tendency to predict longer runs is reflected in T_k values that remain nearly constant for $k \geq 2$. Since *Ss* also predict B to repeat less often for successive blocks of 250 trials, T_k for $k = 0$ also increases with each block of 250 trials. By Trials 751–1000, T_k is constant over k from 0 to 6. The monotonic increase in T_k as a function of k on Trials 251–500 and 501–750 is caused by the gradual increase of T_k for $k = 0$ compared with the rapid increase for $k = 6$. The *S* predicts longer A runs sooner than he ceases to predict repetitions of B. As a result, T_k increases monotonically as a function of k and briefly fulfills the assertion of the conditioning axiom.

DISCUSSION

The results of these experiments show that the assertion of the conditioning axiom that is basic to linear learning models does not fit the fine grain of the data in many types of human binary choice experiments. Variations in temporal parameters, incentive parameters, and the contingency of the events on *S*'s response altered probability matching but did not result in a proportion of A predictions that increased monotonically as a function of the successive occurrence of the event. In every case *Ss* behaved as if they believed the gambler's fallacy, i.e., they did not predict very long or very short runs.

When the experiment was extended to 1000 trials *Ss* began to predict longer runs for successive 250-trial blocks. As a result the proportion of A predictions was the monotonically increasing function of its successive occurrence predicted by the conditioning axiom.

On the final 250 trials *Ss* predicted fewer repetitions of B than in the other blocks of 250 trials. These results agree with those of Goodnow and Pettigrew (1955) and Anderson (1960) who found that *S* will repeat predictions of events that tend to repeat and will not predict repetitions of events that do not tend to repeat. A is predicted to continue to occur or to follow the occurrence of B; B is predicted only after long runs of A.

The monotonically increasing function of T_k against k which is found on Trials 251–750 is a result of the comparatively rapid development of prediction of long runs of A with respect to the decrease in the prediction of repetitions of B. It must be concluded that, in general, the assertion of the conditioning axiom does not give an accurate description of human binary prediction.

SUMMARY

The generality of the conditioning axiom basic to linear learning models was investigated. Ninety *Ss* in groups of 10 predicted a 250-trial 75:25 schedule under various conditions of temporal duration, incentive, contingency, and experience with the schedule. "Restricted latency" instructions plus a criterion of at least 75% of the responses within .5 sec. resulted in probability matching and "contingent-risk" resulted in nearly 100% prediction of the more frequent event (A). However, a sample of the proportion of prediction of A (T_k) as a function of its successive occurrence (k) did not show the monotonically increasing function asserted by the conditioning axiom.

When *S*'s experience with the schedule was extended to 1000 trials, T_k did increase monotonically as a function of k for Trials 251 to 750. The monotonic increase was the result of the differential rate at which *S*s began to predict long runs of A and decreased the prediction of repetitions of B. It was concluded that the assertion of the conditioning axiom does not describe most human binary prediction.

References

Anderson, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

Anderson, N. H., & Whalen, R. E. Likelihood judgments and sequential effects in a two-choice probability learning situation. *J. exp. Psychol.*, 1960, **60**, 111–120.

Bush, R. R., & Mosteller, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323. (a)

Bush, R. R., & Mosteller, F. A mathematical model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423. (b)

Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

Detambel, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.*, 1955, **49**, 97–105.

Edwards, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

Edwards, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Estes, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

Galanter, E. H., & Gerstenhaber, M. On thought: The extrinsic theory. *Psychol. Rev.*, 1956, **63**, 218–227.

Galanter, E. H., & Smith, W. A. S. Some experiments on a simple thought problem. *Amer. J. Psychol.*, 1958, **71**, 359–366.

Goodfellow, L. D. A psychological interpretation of the results of the Zenith Radio experiments in telepathy. *J. exp. Psychol.*, 1938, **23**, 601–632.

Goodnow, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116. (a)

Goodnow, J. J. Response sequences in a pair of two-choice probability situations. *Amer. J. Psychol.*, 1955, **68**, 624–630. (b)

Goodnow, J. J., & Pettigrew, T. F. Effect of prior patterns of experience upon strategies and learning sets. *J. exp. Psychol.*, 1955, **49**, 381–389.

Goodnow, J. J., & Postman, L. Probability learning in a problem-solving situation. *J. exp. Psychol.*, 1955, **49**, 16–22.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Grant, D. A., Hornseth, J. P., & Hake, H. W. The influence of the intertrial interval on the Humphreys' "random reinforcement" effect during the extinction of a verbal response. *J. exp. Psychol.*, 1950, **40**, 609–612.

Hake, H. W., & Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

Humphreys, L. G. Acquisition and extinction of verbal expectation in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

Lindman, H., & Edwards, W. Supplementary report: Unlearning the gambler's fallacy. *J. exp. Psychol.*, 1961, **62**, 630.

Mosteller, F., & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954.

Neimark, E. D. Effects of types of non-reinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

Nicks, D. C. Predictions of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

Rand Corporation. *A million random digits with 100,000 normal deviates.* Glencoe, Ill.: Free Press, 1955.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Siegel, S., & Goldstein, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

Skinner, B. F. The processes involved in the repeated guessing of alternatives. *J. exp. Psychol.*, 1942, **30**, 495–503.

Smith, M. H. Spread of effect is the spurious result of non-random response tendencies. *J. exp. Psychol.*, 1949, **39**, 355–368.

(Received March 4, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 546–554

MULTIDIMENSIONAL SCALING OF FACIAL EXPRESSIONS

ROBERT P. ABELSON AND VELLO SERMAT

Yale University

The ability to identify the emotions displayed in facial expressions has long been a subject of psychological study. The early literature, reviewed by Woodworth (1938), focused attention on the surprisingly frequent inability of untrained *Ss* to report the emotion intended by the stimulus. Woodworth pointed out that errors of identification usually fell into categories psychologically similar to the intended one. A long struggle to organize these similarities in some coherent way culminated in Schlosberg's (1952, 1954) proposals of a two-dimensional and later a three-dimensional theory of facial expressions. In studies by Engen and Levy (1956), Engen, Levy, and Schlosberg (1957, 1958), and Triandis and Lambert (1958) it was shown that judges can reliably locate posed facial expressions along each of Schlosberg's three scales: pleasant-unpleasant, attention-rejection, and sleep-tension. Any given emotional expression can thus be represented by three coordinates, the respective locations on the three component scales. Engen, Levy, and Schlosberg (1958) have given these coordinates for 48 photographs (the "Lightfoot Series") portraying a variety of situations, moods, and emotions via the face of a young actress; e.g., maternal love, anxiety that her plane will crash, savoring a Coca-Cola, etc. The coordinates of each stimulus are the median ratings by 96 male undergraduates on three nine-point scales.

Generally speaking, facial expressions with similar coordinates on Schlosberg's three dimensions are susceptible to confusion with one another, e.g., joy with plasant surprise, hatred with disgust, etc. This elucidates the classical finding of frequent misidentification, and also exemplifies the general relationship between psychological similarity and the probability of "intrusion errors."

However, the particular three dimensions used by Schlosberg are not necessarily the only dimensions or the best dimensions for explaining confusion data or for general usage in this research area. There is the possibility that one or more of Schlosberg's scales, while understandable when made explicit to judges, are unimportant in uninstructed perception of facial expression; or conversely, that one or more important scales have been omitted. The Schlosberg theory thus suffers from the weakness that no direct account is taken of the dimensions operative for the judges; *E* imposes particular dimensions of his own choosing and is arbitrarily forced to give them equal weight.

This vulnerable aspect of the Schlosberg system arises because the historical development of the system was quite indirect. There is only a very loose logical correspondence between confusion data on the one hand and scale locations on the other. A direct correspondence can best be obtained by applying a single analytical operation to a set of confusion data to derive a multidimensional set of scale values. Shepard's (1957) scaling procedure represents one way

to achieve this. An alternative method is detailed below.

A few attempts have already been made to derive scale values directly from confusion or similarity data. Unfortunately, these studies for one reason or another are not directly applicable to the present problem. Hulin and Katz (1935) made an early attempt, but since sophisticated analytical methods were not available at the time, their data were not gathered in the most useful way. Nummenmaa and Kauranne (1958) applied Ekman's (1954) factoring technique to judgments of similarity of facial expressions; however, Ekman's technique spuriously increases the dimensionality of the solution and does not produce clear results, as surmised by Messick and Abelson (1957) and demonstrated empirically by Mellinger (1958). Osgood (1955) produced an interesting three-dimensional facial system mediated by verbal labels attached to poses. The use of verbal mediation and other complicating factors make this study difficult to interpret in the present context.

The multidimensional scaling techniques developed recently by Torgerson (1952) and others permit investigation of a stimulus domain *without preselection of component dimensions*. Judges are asked to respond to stimuli in terms of their similarity to each other. From these judgments, estimates of the "distances" between stimuli are derived. Coordinates of the stimuli in a psychological space of appropriate dimensionality are obtained by a matrix factoring procedure. The identification of the dimensions is post hoc, as in factor analysis.

The present paper reports an application of multidimensional scaling to the domain of facial expressions. Here, the existence of Schlosberg's scales is very useful, for it allows better interpretation of the dimensions derived from the multidimensional scaling operation.

METHOD

Subjects.—The *Ss* were 30 women graduate and special students residing at a women's dormitory in New Haven, Connecticut. Ages ranged from 20–55 yr., with a median of 23 yr. The *Ss* were solicited by posting an advance notice briefly indicating the topic of study and requesting volunteers at designated times. Three groups of 10 *Ss* each were used.

Procedure.—Thirteen photographs had been selected from the Lightfoot Series[1] to tap collectively the full range of each of Schlosberg's three dimensions. All 78 possible pairs of these stimuli were presented to *Ss* in a single random order. The stimulus pairs were projected onto a screen by means of a pair of slide projectors. Each individual stimulus appeared 12 times, 6 times on the left and 6 on the right. Each stimulus pair was exposed for 6 sec., during which time each *S* judged the dissimilarity of the members of the pair and recorded a number from 1 to 9 on a response sheet. The *E* then removed the slides, readied and announced a new pair, and shifted the new slides into position simultaneously. This pause between presentations lasted 10 sec. Pairs were designated on *S*'s response sheets by code letters.

The judgment procedure was preceded by a warm-up period during which the rating scale was explained and three pairs of stimuli were shown as illustrations. The warm-up stimuli were Lightfoot Faces different from the 13 crucial stimuli.

Instructions.—The rating scale instructions to *Ss* were as follows:

> I am going to show you a number of photographs of the same person, and each time I will present two of these pictures side by side. Your task will be to judge how much the two pictures differ in their emotional expression. I would prefer if you try to make a rather spontaneous judgment, without trying to theorize too much as to what emotion is being expressed and how this should relate to the emotional expression of the other picture. Rather you should think in about these terms: emotional expressions of people differ, some being more similar than others. If we should give the difference a numerical score, we would quite arbitrarily decide that we will give the score of 1 to a pair of two facial expressions which are most alike in their emotional content; a score of 9 to two expressions which are furthest apart, and a score of 5 to a pair of pictures which fall about half-way between these two extremes.
>
> Just to give you some idea what I have in mind, I will now show you a few examples. This first pair of pictures portrays

[1] The authors are grateful to Harold Schlosberg for generously providing the necessary slides.

emotional expressions which can be said to be fairly similar; you might score this pair 1 or 2 or 3. . . . The next pair shows facial expressions which are close to the extreme opposites, so you might score them from 7 to 9. . . . This third pair fits in somewhere between the two preceding pairs, so you might score it 4 or 5 or 6. . . . Are there any questions?

Now would you please look at your sheets. There are spaces left for 78 entries, each entry consisting of one score between the values 1 and 9 for one pair of faces. The same pictures will occur again and again, but in different combinations, and you are not rating single pictures but rather the difference between any two pictures at a given time. Each time when I show a pair, I call out two letters, to indicate which entry on the sheet this pair refers to. This is to avoid errors in writing scores into their proper locations on the sheet. Please check each time that the two letters I named are exactly the same as the letters before the empty space where you are going to write in your score for this pair.

Please try to write in your judgments as quickly as possible; thinking too long will only confuse you. It is also absolutely essential that you do not miss *any* single entry; even if it is difficult to judge some pictures, put down your first best guess. Leaving some entries empty will create much more serious problems for my re-research than being uncertain about the correct score and putting in the best guess. Also, try to use the whole scale of numbers, that is, use every number from 1 to 9 at some time or other. And do not go back to check what you have written down before. Just go on from one pair to another. And of course, please avoid influencing one another, or discussing, or exchanging any opinions or advices throughout this time.

Are there any further questions? Fine, then we will start with the first pair of facial expressions. Please look at it and then put down your best estimate by writing a number into the first entry space.

Remember: rating 1 indicates that the two faces differ very, very little in their emotional expression, rating 9 indicates that they differ extremely much, and the other numbers represent gradations of difference in between.

The entire judgment session lasted approximately 30 min. The data consisted of the dissimilarity ratings by 30 *S*s of 78 stimulus pairs. The data were later punched into IBM cards for the subsequent analysis.

Results

Table 1 reproduces from Engen, Levy, and Schlosberg (1958) the scale values of the 13 stimuli on the three Schlosberg scales. One notes in this table a strong positive correlation between the values on Dimensions

TABLE 1

Scale Values on the Three Schlosberg Dimensions for 13 Faces from the Lightfoot Series

Stimulus	Pose	(V_1) Pleasant-Unpleasant	(V_2) Attention-Rejection	(V_3) Tension-Sleep
7	Grief at death of mother	3.8	4.2	4.1
13	Savoring a Coke	5.9	5.4	4.8
15	Very pleasant surprise	8.8	7.8	7.1
16	Maternal love—baby in arms	7.0	5.9	4.0
20	Physical exhaustion	3.3	2.5	3.1
28	Anxiety—something is wrong with her plane	3.5	6.1	6.8[a]
29	Anger at seeing dog beaten	2.1	8.0	8.2
30	Physical strain—pulling hard on seat of chair	6.7	4.2	6.6
32	Unexpectedly meets old boy friend	7.4	6.8	5.9
36	Revulsion	2.9	3.0	5.1
37	Extreme pain	2.2	2.2	6.4
51	Knows her plane will crash	1.1	8.6	8.9
56	Light sleep	4.1	1.3	1.0

[a] The entry 8.2 in the original published table was a misprint.

TABLE 2

SQUARED DISTANCES, d^2, BETWEEN STIMULI, AS DERIVED FROM DISSIMILARITY RATINGS

Stimulus	7	13	15	16	20	28	29	30	32	36	37	51	56
7	0.00	4.05	8.25	5.57	1.15	2.97	4.34	4.90	6.25	1.55	1.68	6.57	3.93
13	4.05	0.00	2.54	2.69	2.67	3.88	8.53	1.31	1.88	4.84	5.81	7.43	4.51
15	8.25	2.54	0.00	2.11	8.98	9.27	11.87	2.56	0.74	9.25	7.92	8.30	8.47
16	5.57	2.69	2.11	0.00	3.78	6.05	9.78	4.21	0.45	4.92	5.42	8.93	3.48
20	1.15	2.67	8.98	3.78	0.00	2.34	7.12	5.90	4.77	2.22	4.34	8.16	1.60
28	2.97	3.88	9.27	6.05	2.34	0.00	1.36	5.18	5.45	4.17	4.72	4.66	4.89
29	4.34	8.53	11.87	9.78	7.12	1.36	0.00	8.47	10.20	5.44	4.31	1.57	9.18
30	4.90	1.31	2.56	4.21	5.90	5.18	8.47	0.00	2.63	5.45	3.79	6.49	6.05
32	6.25	1.88	0.74	0.45	4.77	5.45	10.20	2.63	0.00	7.10	6.58	9.77	6.55
36	1.55	4.84	9.25	4.92	2.22	4.17	5.44	5.45	7.10	0.00	1.98	4.93	4.12
37	1.68	5.81	7.92	5.42	4.34	4.72	4.31	3.79	6.58	1.98	0.00	4.83	3.51
51	6.57	7.43	8.30	8.93	8.16	4.66	1.57	6.49	9.77	4.93	4.83	0.00	12.65
56	3.93	4.51	8.47	3.48	1.60	4.89	9.18	6.05	6.55	4.12	3.51	12.65	0.00

V_2 and V_3. This is not an accident of selection for the present study of 13 faces out of the total set of 48, since the correlation between these two dimensions in the total set is even higher.

The configuration of stimuli in Schlosberg's three-coordinate space is therefore quite flat; i.e., a plane is not a bad fit to the stimulus points. This fact becomes important later.

For each stimulus pair, a distribution of dissimilarity ratings was available from the present study. The mean of this distribution is a possible estimate of the "distance" between the members of the pair; however, a more sensitive estimate can be obtained from the method of successive intervals. Diederich, Messick, and Tucker's (1957) version of this method was used,[2] yielding estimates of the distances between each stimulus and every other. These distances are determined on an interval scale only; a zero point on the distance scale was located by Messick and Abelson's (1956) procedure. Table 2 presents the *squares* of the resulting distances between each stimulus and every other. The self-distance of each stimulus is zero.

Before proceeding to the multidimensional analysis of the distance matrix, it is possible to make comparisons between the present distance data and the previous scale data. Suppose one assumes that the interstimulus distances can be explained by interstimulus differences on the Schlosberg scales; i.e., that a three-dimensional space composed of the Schlosberg dimensions locates the stimuli so as to generate the distance data.

In setting up such a comparative analysis, the following considerations were employed:

1. Euclidean geometry requires that the squared distance between Stimuli j and k equal the sum of the squared scale differences between Stimuli j and k on each of the three component dimensions.

2. However, since the ranges of the Schlosberg scales are not calibrated one with another, the three component dimensions might have to be weighted differentially.

3. One must realistically expect the interstimulus distances to depend at least to a minor extent upon a variety of "extraneous"

[2] Carl Helm of the Educational Testing Service graciously performed this analysis.

dimensions. Each such dimension will *increase* the interstimulus distances beyond what one would predict from the Schlosberg scales. This variable error can be partitioned conceptually into a constant error and a random error. The latter category also serves to include sampling and measurement errors in the distance data.

4. The Schlosberg scale values are assumed to have been measured without error.

The form of the assumed relationship is thus:

$$d_{jk}^2 = b_1(V_{1j} - V_{1k})^2 + b_2(V_{2j} - V_{2k})^2 + b_3(V_{3j} - V_{3k})^2 + C + e_{jk} \quad [1]$$

for all $j \neq k$, where the d_{jk}^2 are the distance data of Table 2, the quantities $(V_{1j} - V_{1k})^2$, etc. are computed from the scale data of Table 1, the b's and C are unknown constants, and the e_{jk} are errors with zero mean and a "random" distribution.

The form of the problem is that of a multiple regression with an N of 78 stimulus pairs, each pair providing a criterion observation d_{jk}^2 and three predictor observations of the form $(V_j - V_k)^2$. The unknown constants were determined by the usual least squares procedure, and the results are presented in the first column of Table 3.

The very high multiple R associated with the regression indicates that the Schlosberg scales predict three-quarters of the variance in the distance data. The regression weights suggest a particular calibration of the three scales (an unachieved goal often mentioned in the literature). It follows from Equation 1 that if the scales are to have equal weight in predicting interstimulus distance the *values* on each scale must be multiplied by a factor proportional to the square root of the appropriate b. A rough indication of these factors will suffice, since the present modest collection of data is not the last word in the matter. Using all three scales, a rough calibration is provided by the factors 4:3:2. In other words, if pleasant-unpleasant is located on a scale from 1 to 9, then tension-sleep should be located on a scale from 1 to 7, and attention-rejection on a scale from 1 to 5.

TABLE 3

WEIGHTS TO APPLY TO THE SCHLOSBERG SCALES TO BEST APPROXIMATE THE DISTANCE DATA ACCORDING TO EQUATION 1

Scale	V_1, V_2, & V_3	V_1 & V_3	V_1 & V_2	Best Single Scale
V_1	.156	.157	.148	.151
V_2	.034	—	.088	—
V_3	.091	.112	—	—
Constant error, C	1.72	2.16	2.53	3.49
Multiple R	.868	.854	.811	.706

Table 3 also gives the results for two-scale regressions, omitting one scale from the predictive formula. Omission of attention-rejection does not lower prediction appreciably; omission of tension-sleep is slightly more damaging. The reason attention-rejection is less helpful in prediction is not because it is a poor scale, but rather because of redundancy: squared scale differences on this scale correlate *positively* with squared scale differences on pleasant-unpleasant, whereas squared scale differences on tension-sleep correlate *negatively* with squared scale differences on pleasant-unpleasant.

Considering only two scales at a time, the rough calibration factors are: for V_1 and V_3, 5:4; for V_1 and V_2, 4:3.

The last column of Table 3 indicates that the best single contribution to prediction is, not surprisingly, made by the pleasant-unpleasant dimension, which accounts for half the variance in the distance data.

The next step was to derive a multidimensional set of stimulus coordinates from the distance data with-

TABLE 4

MULTIDIMENSIONAL SCALING RESULTS: COORDINATES OF THE STIMULI ON FIVE DIMENSIONS

Stimulus	Dimension				
	I	II	III	IV	V
7	−.65	−.64	.24	−.31	−.24
13	.85	.06	−.36	−.71	−.26
15	1.64	1.02	.22	.19	.06
16	1.15	−.16	−.05	.87	−.14
20	−.19	−1.14	−.48	−.11	−.51
28	−.84	−.01	−1.09	−.17	.28
29	−1.74	.77	−.48	.30	.43
30	.73	.54	.42	−.90	.37
32	1.43	.32	−.40	.26	−.04
36	−.72	−.55	.65	.10	−.71
37	−.66	−.23	1.05	.10	.51
51	−1.23	1.55	.21	.16	−.39
56	.22	−1.52	.06	.23	.64
Sum of squares	14.02	8.83	3.80	2.50	2.14

out reference to the Schlosberg scales. The details of the standard procedure are given by Torgerson (1958).

The dimensionality of the solution is determined from the pattern of latent roots of a Matrix B derived from the matrix of squared distances. For the present data, five dimensions were required. The coordinates of the stimuli on these five orthogonal dimensions are given in Table 4. The row labeled "sum of squares" gives the sum of squared distances of stimulus points from the origin on each dimension. This quantity may be thought of as the variance or "spread" accounted for by each dimension. It is seen that the first two dimensions are far more important than the latter three, accounting between them for 73% of the total spread. The factoring procedure could have been stopped after two dimensions to provide a very crude solution.

The dimensions are initially unidentified. As in factor analysis, a rotation of the coordinate axes may be necessary to yield psychologically interpretable dimensions, although the "simple structure" criterion of factor analytic rotation is not necessarily relevant. In the present case, the Schlosberg scales provide useful guidance in interpretation of dimensions. Table 5 gives the correlations between coordinates on the five present dimensions and values on the three Schlosberg scales. It should be remembered that two different samples of judges were involved, differing in age and sex. The judgment task itself was different in that no reference was ever made in the present task to any explicit dimensions.

In view of this, the correlations of .947 between Dimension I and pleasant-unpleasant, .878 between Dimension II and attention-rejection, and .917 between Dimension II and tension-sleep are extremely high. The multiple correlations, .962, .982, and .965, given in the last column of Table 5 show that the each of the Schlosberg scales can be located without appreciable error by particular orientations of the Axes I–V. None of the Schlosberg scales, therefore, can be called trivial in the sense that the stimulus separations it provides are not to be found in direct judgments of dissimilarity. However, one again encounters redundancy between Scales V_2 and V_3; Dimension II is representative of both attention-rejection and tension-sleep. Thus while there are

TABLE 5

CORRELATIONS BETWEEN THE PRESENT FIVE DIMENSIONS AND THE THREE SCHLOSBERG SCALES

Schlosberg Scale	Dimension					Multiple R (all five)
	I	II	III	IV	V	
V_1	.947	.121	−.094	−.031	.061	.962
V_2	.026	.878	−.374	.210	.094	.982
V_3	−.295	.917	.011	−.041	.046	.965

three scales and five dimensions, there are essentially three extra, uninterpreted dimensions rather than two.

Attempts were made to give psychological identification to the three extra dimensions. One may seek rotations of Axes III, IV, and V of Table 4 yielding meaningful interpretations. Alternatively, a somewhat more sophisticated approach would be as follows: (*a*) locate a new axis I′ correlating maximally with V_1; (*b*) locate a new axis II′ correlating maximally with an optimal combination of V_2 and V_3; (*c*) solve for a triplet of axes, III′, IV′, and V′ orthogonal to I′ and II′; and (*d*) rotate III′, IV′, and V′ to meaningful positions. Both approaches were tried, but neither met with sufficient success to justify pronouncements on the nature of the extra dimensions. The subjective verdict was that these dimensions were "hash."

Discussion

The present study provides evidence that the Schlosberg scales are sound: they predict the dissimilarity data fairly well, and the extra dimensions that appear in the multidimensional scaling operation do not suggest new meaningful scales to replace or supplement the old ones.

The Schlosberg scales do suffer, however, from redundancy in that two scales seem to serve as well as three. If a choice must be made between tension-sleep and attention-rejection, the former deserves a very slight preference. As a supplement to the pleasant-unpleasant scale, the tension-sleep scale is better than the attention-rejection scale in predicting the dissimilarity data. Other relative demerits of the attention-rejection scale have been noted elsewhere: difficulty in explaining the scale to *Ss* (Schlosberg, 1952); lower inter- and intrarater reliability (Engen et al., 1958); lower stability across scaling methods (Engen & Levy, 1956). There is of course the possibility of devising a new scale which compromises between the two scales in question.

Osgood (1960, p. 362) reports a study done with Hastorf in which facial expressions were rated on scales of *pleasantness* and *activation*. This two-dimensional scheme, apparently a selective refinement from Osgood's (1955) earlier three-factor system, is probably close to optimal. The earlier three-dimensional system is hard to interpret, since it is based on different poses by different actors, thus confounding inter- and intraindividual dimensions of expression.

There are several possible explanations for the uninterpretibility of the three extra dimensions of the present study. Conceivably, sensible interpretations were overlooked. Alternatively, it may be that the emergence of these dimensions was somehow an artifact of the method of dealing with the distance data.

There are three possibilities here: that the assumption of a Euclidean geometry is to some extent inappropriate; that there is a "communality problem," i.e., that each stimulus introduces a small dimension of its own; that some monotone transformation of the distance scale would produce cleaner results. There is little definitive knowledge of how to attack the first two problems, but on the third problem, it has been suggested by Helm, Messick, and Tucker (1959) that an exponential transformation of the distance scale can often reduce dimensionality. Such a transformation was attempted with the present data, but the outcome was unsatisfactory. Instead of reducing the dimensionality, the revised solution additionally produced a sizeable negative root in the factoring process, i.e., an imaginary dimension. Other transformations which stretch large distances would have the same effect. Another possible source of extra, uninterpretable dimensions might lie in individual differences between *Ss* in the manner of judging dissimilarity. Beyond the two big dimensions, a great variety of criteria can be used to judge additional distance be-

tween stimuli. If different *S*s use different criteria, the general result may well be "hash." As a final explanation or rationalization, we submit that it is perhaps too extravagant an expectation that minor dimensions should have a ready interpretation with such complex stimuli as facial expressions. At all events, the first two dimensions emerge strongly and clearly, embodying the important outcome of this study.

A persistent concern in the facial expressions area has been the shape of the stimulus surface (the "emotion solid"). Schlosberg (1952) designated the shape as roughly oval in two dimensions and (Schlosberg, 1954) as a tilted cone with an oval cross-section in three dimensions. Triandis and Lambert (1958) strove inconclusively to modify Schlosberg's formulation. One difficulty besetting these attempts has been the lack of calibration between dimensions.

Shape is a direct outcome of the multidimensional scaling operation. A plot of Dimensions I and II of the present study yields a stimulus boundary or envelope which is nearly an equilateral triangle. The "sleep" expression, Stimulus 56, anchors one corner of the triangle at the negative pole of Dimension II. The other two corners lie at the other pole of Dimension II such that the side joining them is parallel to the I axis; Stimulus 15, very pleasant surprise, is close to one corner and Stimuli 29 and 51, anger and fear, close to the other. The boundary could of course bulge were more stimuli added, with a coat-of-arms or pear-shaped contour as plausible possible alternatives. An oval or elliptical shape appears implausible, although one cannot have full confidence in this conclusion without a much more exhaustive sampling of stimuli.

A final word of clarification is in order. This study and all others like it can only achieve a representation of facial expressions *as interpreted by the viewer*. There is sometimes an implication in studies of this type that one might somehow identify the full range of possibilities for the "sender," i.e., discover the "true" dimensions of emotion. It is unwarranted to expect to complete this ambitious task by procedures involving viewers' judgments. Nuance is blurred or lost through misinterpretation, though this loss is not nearly so complete as was supposed in the earliest studies.

Summary

Thirty women gave nine-point rating scale judgments of the dissimilarity of emotional expression of each of 78 pairs of facial poses formed from combinations of 13 diverse stimuli of the Lightfoot Series. The dissimilarity ratings were converted to interstimulus distances. Comparison of these distances with the distances one would expect on the basis of the three Schlosberg scales of facial expression yielded the following conclusions: (*a*) The pleasant-unpleasant scale explains 50% of the variance in the distance data; the pleasant-unpleasant (V_1) and tension-sleep (V_3) scales together account for almost 75% of the variance; the attention-rejection (V_2) adds little more to prediction. (*b*) If the V_2 scale is omitted, the V_1 and V_3 scales should be calibrated in approximately the ratio 5:4. (*c*) Standard multidimensional factoring techniques applied to the distance matrix yielded stimulus coordinates on five dimensions, with the major dimension essentially identical to Schlosberg's pleasant-unpleasant scale, the second dimension, a compromise between Schlosberg's other two scales, correlating extremely highly with both, and the remaining three dimensions not readily interpretable. (*d*) The shape of the configuration of stimuli in the space of the first two dimensions is approximately an equilateral triangle with a sleep stimulus at one corner and actively pleasant and unpleasant expressions at the other two corners. (*e*) If the Lightfoot Series has indeed captured the full range of potential expressions, then for usage a two-dimensional system would appear to be adequate: pleasant-unpleasant and tension-sleep, or perhaps a slight variant of the latter.

References

Diederich, G., Messick, S. J., & Tucker, L. R. A general least squares solution for successive intervals. *Psychometrika*, 1957, **22**, 159–173.

Ekman, G. Dimensions of color vision. *J. Psychol.*, 1954, **38**, 467–474.

ENGEN, T., & LEVY, N. Constant-sum judgments of facial expressions. *J. exp. Psychol.*, 1956, **51,** 396–398.

ENGEN, T., LEVY, N., & SCHLOSBERG, H. A new series of facial expressions. *Amer. Psychologist*, 1957, **12,** 264–266.

ENGEN, T., LEVY, N., & SCHLOSBERG, H. The dimensional analysis of a new series of facial expressions. *J. exp. Psychol.*, 1958, **55,** 454–458.

HELM, C., MESSICK, S. J., & TUCKER, L. R. Psychophysical law and scaling modes. Princeton: *Educ. Testing Serv. res. Bull.*, 1959, No. RB-59-1.

HULIN, W. S., & KATZ, D. The Frois-Wittman pictures of facial expressions. *J. exp. Psychol.*, 1935, **18,** 482–498.

MELLINGER, J. J. A comparison of multidimensional scaling and similarity analysis. *Amer. Psychologist*, 1958, **13,** 375. (Abstract)

MESSICK, S. J., & ABELSON, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, **21,** 1–16.

MESSICK, S. J., & ABELSON, R. P. Scaling and measurement theory. *Rev. educ. Res.*, 1957, **27,** 487–497.

NUMMENMAA, T., & KAURANNE, U. Dimensions of facial expression. *Acad. Paedag. Jyvaskylaensis res. Rep.*, 1958, No. 20.

OSGOOD, C. E. Fidelity and reliability. In H. Quaster (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955.

OSGOOD, C. E. Cognitive dynamics in human affairs. *Public opin. Quart.*, 1960, **24,** 341–365.

SCHLOSBERG, H. The description of facial expressions in terms of two dimensions. *J. exp. Psychol.*, 1952, **44,** 229–237.

SCHLOSBERG, H. Three dimensions of emodimensions of emotion. *Psychol. Rev.*, 1954, **61,** 81–88.

SHEPARD, R. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22,** 325–346.

TORGERSON, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika*, 1952, **17,** 401–419.

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

TRIANDIS, H. C., & LAMBERT, W. H. A restatement and test of Schlosberg's theory of emotion with two kinds of subjects from Greece. *J. abnorm. soc. Psychol.*, 1958, **56,** 321–328.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received March 25, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 555–562

SPONTANEOUS RECOVERY INTERVAL AS A FACTOR IN REACQUISITION OF T MAZE BEHAVIOR [1]

JOHN W. COTTON
University of California, Santa Barbara

GLEN D. JENSEN
University of New Mexico

AND DONALD J. LEWIS
Rutgers University

The present study attempts to show the effects of spontaneous recovery intervals ranging from 1 min. to 9 days upon reacquisition of both speed and correct turns in a T maze. This is a wider range of intervals than previously studied. Further, most available data are from other apparatuses. Data from the first reacquisition trial may be taken as an indication of spontaneous recovery, traditionally defined. Later reacquisition trials are studied as dependent variables of interest in their own right. A second purpose of this experiment is to provide definitive empirical curves of acquisition and extinction behavior in a T maze. For that reason experimental conditions are held constant for all animals during acquisition and extinction.

METHOD

Subjects.—The *S*s were 100 naive male albino rats purchased from the Sprague-Dawley Company of Madison, Wisconsin. An additional 17 animals were discarded, 7 because of sickness or death, 8 because of apparatus failure or procedural error, and 2 because of failure to learn the running response. The *S*s were from 55 to 80 days old at the beginning of experimentation, with a mean of 63.1 days. They were assigned at random to five groups, subject to the restrictions that no group have more than 1 more *S* completed at any time than any other group and that the total group size be 20 at the end of the experiment.

Apparatus.—The "narrow" T maze (2 in. inside width) used by Cotton, Lewis, and Metzger (1958) was employed. An automatic timing device measured the running time in .01-sec. units from the start box to a point 46¼ in. down the stem and a second timer began with the first and stopped when *S* reached the entrance to one goal box or the other. The readings of the second timer are not reported in this study. They were used as reported in the extinction procedure.

Procedure.—Six stages constituted the procedure. The period of deprivation and taming extended for 7 days in which each *S* was under 23-hr. food deprivation with 1 hr. of ad lib. feeding and a 5-min. period of handling each day. Each *S* was allowed to eat up to six 45-mg. lab chow pellets from a food cup on a table during every handling session. If *S* ate less than six during that session, the remainder were placed in the home cage immediately afterward.

A pretraining session on Day 8 consisted of 10 rewarded trials through the complete T maze for each *S* with a 15-sec. rest interval in the home cage between the end of one trial and the beginning of the next. Odd-numbered trials were free-choice trials; even-numbered trials were forced to the side opposite the preceding response. The deprivation period at the beginning of this and all subsequent sessions was 22 hr. except when stated otherwise below. One 45-mg. lab chow pellet was given on all rewarded trials following Day 7.

An acquisition session on Day 9 consisted of 30 noncorrection trials through the complete T maze. For each *S* the side less frequently chosen on free-choice trials of Day 8 was designated the correct side on Day 9. Turns to the correct side were then rewarded, and turns to the incorrect side were not rewarded. The rest intervals in this and succeeding periods were as on Day 8.

An extinction session on Day 10 consisted of 30 unrewarded trials through the complete T maze with the exception that, if *S* failed to

[1] This experiment was performed at Northwestern University under National Foundation Grant G-2672. The preparation of the article itself was supported by National Science Foundation Grant G-8706.

stop the second timer within 2 min., he was removed from the maze. He then received a running time listing of 2 min. unless he had reached the microswitch stopping the first clock. Forty-three such removals were made during extinction but only 7 of these were before *S* stopped the first timer.

The spontaneous recovery period was a delay in the home cage between extinction and reacquisition. Members of the five groups were delayed 1 min., 10 min., 3 hr., 1 day, and 9 days, respectively. Groups 1D and 9D were placed on a 22-hr. food deprivation schedule until reacquisition began; Groups 1M and 10M had a deprivation period of approximately 23 hr. because the extinction session had added 1 hr. to their previous 22-hr. deprivation; similarly, Group 3H was under a deprivation of approximately 26 hr. This variation of 4 hr. in deprivation level may be expected to produce slightly faster running for Group 3H than otherwise. While the range from 22 to 26 hr. of food deprivation has not been explored, Yamaguchi's (1951) study of the range from 3 to 72 hr. suggests that the effect of variation from 22 to 26 hr. of deprivation is negligible.

A reacquisition period at the end of the spontaneous recovery period was identical for each *S* to the acquisition procedure.

Results and Discussion

Comparability of groups.—Mean reciprocal running times were computed for each *S* for each stage of learning. Analyses of variance were conducted upon the means of the last 6 acquisition trials, the means of the 30 extinction trials, and the means of the last 6 extinction trials. None of these five analyses showed significant ($P < .05$) differences among groups, which is as expected since no differen-

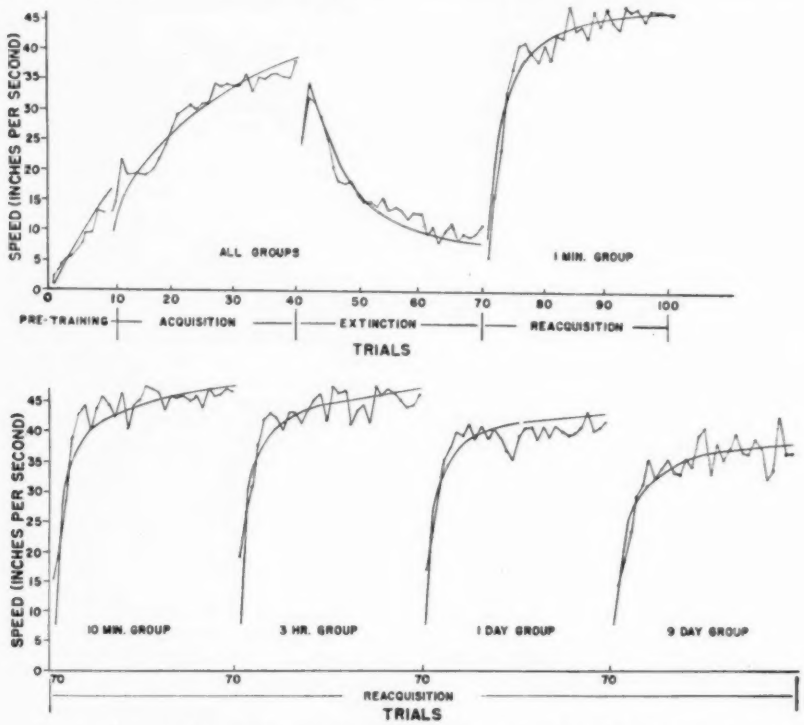

FIG. 1. A comparison of theoretical and empirical mean speeds, with data pooled for groups treated alike.

tial treatment was introduced until the spontaneous recovery period. The analysis of the last 6 acquisition trials gave a hint of initial differences among groups, with an F at the .07 level suggesting systematically slower running times by Group 9D.

Analyses of variance of numbers of correct responses for the same trial blocks just discussed also failed to show significant differences among groups. It is concluded that the group assignment procedure led to groups which may be considered comparable at the beginning of the retention period.

Effect of recovery interval.—The course of pretraining, acquisition, and extinction for all groups combined, and reacquisition of running speed for each group is shown in Fig. 1, together with fitted curves described later. Group 9D is worst in reacquisition, as it had been in pretraining and acquisition. Group 1M is also poor early in reacquisition but reaches the same level as Groups 10M and 3H, suggesting that minimal recovery over a 1-min. interval does not interfere with subsequent rewarded relearning.

The mean reciprocal running times for the first reacquisition trial were assessed by analyses of variance and covariance, using the mean reciprocal running time for each S on the last 6 extinction trials as the predictor variable. Table 1 summarizes these two analyses, showing that the five groups differ significantly on the first reacquisition trial regardless of whether or not covariance analysis is employed. Because the $r_{\text{within groups}}$ is only .28, the covariance analysis made little change in the F for group differences and no more than .01 difference of any adjusted mean speed from the observed mean. Table 1 also presents the observed and adjusted mean speeds (reciprocal running times multiplied by 46.25, the distance run while timed) for each group. The trend in each case is for substantial spontaneous recovery from Groups 1M to 10M, lesser recovery from Groups 10M to 3H, and then a decline in recovery for Groups 1D and 9D. Since the means for Groups 1D and 9D remain higher than that for Group 1M, these groups show a residual spontaneous recovery. The reduction in speeds for Groups 1D and 9D could indicate forgetting of the running habit. However, since a Tukey simultaneous interval test (Scheffé, 1959, pp. 73–75) shows no significant differences between groups except for the comparison of Groups 1M and 3H, it is more likely a random effect. Note that the four greatest delay groups show a mean increase of 70% or more above the observed mean speed of Group 1M.

TABLE 1

ANALYSES OF VARIANCE AND COVARIANCE OF RECIPROCAL RUNNING TIMES ON FIRST REACQUISITION TRIAL TOGETHER WITH OBSERVED AND ADJUSTED MEAN SPEEDS

Source	Variance Analysis			Covariance Analysis		
	df	*MS*	*F*	*df*	*MS*	*F*
Among groups	4	.193	3.06*	4	.168	2.87*
Within groups	95	.063	—	94	.058	—

Group	Observed Mean Speed (In/Sec)	Adjusted Mean Speed (In/Sec)
1 min.	9.71	10.13
10 min.	18.68	18.78
3 hr.	21.46	21.00
1 day	19.75	19.56
9 days	16.97	17.16

* $P < .05$.

Analyses of variance and covariance of mean speeds for all 30 reacquisition trials, using speeds from the last 6 extinction trials as the predictor variable, again both showed

significant group differences. These analyses are not tabled here because of their similarity to the contents of Table 1. Both observed and adjusted mean speeds increase from Group 1M to Group 10M and decrease thereafter. No increment in speed compared with Group 1M appears beyond the 3-hr. condition. The mean for Group 9D is 5.26 in/sec below that of any other group. Analyses of variance and covariance omitting Group 9D yield moderate but nonsignificant Fs ($P > .05$ in each case).

The mean percentages of correct responses for all groups combined in acquisition and extinction and for each group separately during reacquisition appear in Fig. 2, together with fitted curves to be described later. Percentages are not graphed for pretraining because even-numbered trials were always forced, and no differential reinforcement had yet been instituted. In early reacquisition trials Groups 1M and 9D appear inferior. No suggestion of group differences prior to the reacquisition period was found.

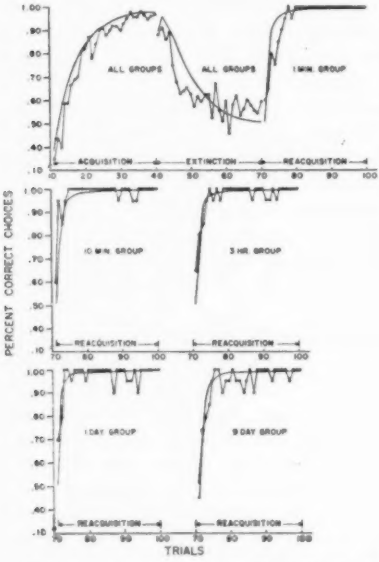

FIG. 2. A comparison of theoretical and empirical percentages of correct choices, with data pooled for groups treated alike.

Statistical analysis of turn data from the first reacquisition trial gives no indication of recovery interval effects. The mean percentage of correct responses for all groups combined is 60 on that trial, exactly the same as on the last extinction trial. The percentages of Ss making the correct response on the first reacquisition trial are 60, 60, 65, 70, and 45 for the five groups arranged in order from 1M to 9D. The χ^2 obtained in a test for differences among these five percentages is 2.92, far below the 9.5 required for significance at the 5% level with $df = 4$. Similarly, a median test comparing the number of Ss per group having more than the median number of correct responses in all 30 reacquisition trials yields a χ^2 of 7.55 ($P > .05$, $df = 4$). The trend of these numbers above the median is, however, suggestive: 11, 8, 8, 6, and 3, with the groups arranged in the same order as above.

This last indication that reacquisition of the correct turn may be retarded with increasing recovery time is bolstered by a tally of the number of Ss having all perfect turns following their first correct choice during reacquisition: These numbers are 17, 14, 11, 10, and 6, with the groups arranged in the same order as before. A 5×2 homogeneity test using these frequencies and the corresponding numbers of Ss not having all perfect trials following one reinforcement yields a χ^2 of 14.6 ($P < .01$, $df = 4$).

Curve-fits of the data.—The presence of responses for 100 Ss per trial, with every S treated identically throughout pretraining, acquisition, and extinction, makes these data useful estimates of population trial by trial means. Accordingly the mean running speeds

and percentages of correct choice have been fitted, using empirical procedures subject to the restriction that Hullian concepts be applied whenever possible and parsimonious.

The population mean running speeds (R_S) are hypothesized to conform to the following basic equation throughout the study:

$$R_S = A + KHR \qquad [1]$$

where A is the lowest speed obtained or to be expected in a given stage of the study, K is Hull's (1951) incentive motivation concept, altered according to Dufort and Kimble's (1956) suggestion to increase gradually during acquisition or following other increases in reinforcement magnitude and to decrease gradually during extinction or other decreases in reinforcement magnitude. As a consesequence K mediates transition phases in studies like those of Crespi (1942) and mediates extinction and reacquisition as well. Inhibition constructs have been omitted because K handles the extinction phenomena of this study and because of the vagueness and inconsistency of Hull's inhibition postulates (Cotton, 1955). The quantity H is Hull's (1943) habit strength concept, subjected to Spence's (1956) contiguity assumption that H increases whenever the response in question is made, rather than only when it is overtly rewarded. The parameter R (for retention percentage) is an expression introduced for the first time by this article. It is intended to reflect the well-known phenomenon (Estes, 1955) that first-trial running speeds (a day after rewarded trials) are depressed compared with speeds at the end of the previous day. The value of R is presumed unity except on Trial 1 of such days.

In a T maze study we may properly think of two K values, one for each side. These will be called K_C, for the side correct during acquisition, and K_I, for the side incorrect during acquisition. It seems acceptable to assume that R_S will be controlled in part by the greater of the two Ks whereas R_P, the probability of the correct response, will be controlled in part by the difference $(K_C - K_I)$ between them. Thus K of Equation 1 becomes:

$$K = \max\,(K_C, K_I) \qquad [2]$$

where:

$$K_C = (K_O - A_P)10^{-iN_S} \qquad [3a]$$

during pretraining or acquisition and:

$$K_C = K'(10^{-kN_S}) \qquad [3b]$$

during extinction (where K' denotes the K_C extrapolated from acquisition theory for the first trial of extinction but corrected for the shift in minimum speeds from pretraining to extinction; i.e.:

$$K' = \frac{K_C H + A_P - A_E}{H} \qquad [4]$$

with K_C and H being extrapolations from acquisition theory) and:

$$K_C = (K_O - A_R)(10^{-jN_S}) \qquad [3c]$$

during reacquisition.

The equation for K_I has the same form as Equation 3a (except for different N_S values) during pretraining but, because it is never rewarded thereafter, takes the form of Equation 3b for the remainder of the experiment. Since there is no shift in minimum speeds from pretraining to acquisition, the analogue of Equation 4 is much simplified.

Previously undefined quantities in the equations above include a rate of growth constant j for acquisition of K, a rate of growth constant k for extinction of K, a maximum value K_O for K based on the magnitude of reinforcement used in the study, and N_S, the number of previous trials in a rewarded series or the number of previous trials in an unrewarded series. All but the last quantity are estimated wholly from the data. Specification of N_S values depended on several considerations: The "correct" side in the present study was the nonpreferred side during pretraining. Because only 20% of free choices during pretraining were to the later correct side and because, despite the fact that forced choices during pretraining equalized experience on both sides, a strong position preference existed at the beginning of acquisition, it will be assumed that each of the 10 pretraining trials added an increment of .2 to N_S for the correct side and .8 to N_S for the incorrect side, yielding values of 2 and 8, respectively, at the end of pretraining. During acquisition, N_S for the correct side increased by 1.0 per trial whereas N_S for the incorrect side started over at 0 and increased by 1.0 per trial because that side was being extinguished.

The two other terms of Equation 1 are:

$$H = 7 = B(10^{-iN}) \qquad [5]$$

where B is a curve-fitting constant, i is a curve-fitting constant showing the rate of growth of H, and N is the number of previous trials of any kind and:

$$R = 1 - C(10^{-mD}) \qquad [6]$$

on Trial 1 of a day following rewarded trials; $=1$ otherwise, where C is a curve-fitting constant, m is a curve-fitting growth constant, and D is the number of previous days of experimentation minus 1.

Figure 1, already shown, presents the theoretical running speed curves obtained

TABLE 2

CURVE-FITTING CONSTANTS USED WITH THE FIVE SPONTANEOUS RECOVERY INTERVALS

Groups	Reacquisition	
	K_O	j
1 min.	48.10	.618
10 min.	49.02	.596
3 hr.	48.33	.513
1 day	43.48	.430
9 days	40.14	.678

All Groups (Pretraining and Acquisition)

B	i	A_P	K_O	j	A_R	k	A_R	C	m
.923	.01725	5.09	47.73	.499	11.33	.0672	11.66	.418	.0561

Note.—A_R is a consequence of the extinction theory, rather than a parameter. It is presented to simplify verification of the theoretical values in Fig. 1.

with Equations 1 through 6, together with the empirical curves on which they were based. Table 2 shows the parameter values used in Fig. 3. All these values are based upon the data of this study, except for C and m which were obtained from unpublished curve-fits of first trial decrements in 9 successive days of 100% reinforcement training (Group A of Jensen & Cotton, 1960). Table 2 shows that K_O decreases systematically on either side of Group 10M, making 10 min. the point of peak recovery as expected from the statistical analyses for all 30 reacquisition trials. The j parameter increases on either side of Group 1D, indicating that rate of reacquisition is greatest for either short or long recovery intervals.

The curve-fitting of the turn data from this experiment is wholly a consequence of the theory presented for speeds, so that the only way in which it depends on turn data is in the previous assumptions about N_S during pretraining. The theory for turn data is consistent with Hull's (1943) Postulates 12 and 16 and (except for the lower limit of integration) with Spence's (1956, p. 203) equation. We assume that momentary effective reaction potential ($\dot{E}$) is normally distributed (except for truncation) with mean $K_C H$ or $K_I H$ depending on the choice in question and with variance σ^2. Hull's Postulate 16 tells us that the correct choice will be made whenever $\dot{E}_C > \dot{E}_I$. It is readily proved that, if we assume truncation of the $\dot{E}_C - \dot{E}_I$ distribution at ± 2.5 from its mean,[2] the probability of a correct response will be 81/80 of the normal curve integral from $z = -2.5$ to:

$$\frac{(K_C - K_I)HR}{\sqrt{2}\sigma}$$

where:

$$z = \frac{\dot{E}_C - \dot{E}_I - (K_C - K_I)HR}{\sqrt{2}\sigma}$$

These values have been computed for every trial in this experiment, and were previously shown in Fig. 2, together with the obtained relative frequencies of correct choice. The quantity $\frac{1}{\sqrt{2}\sigma}$ is determined by the parameters of Equation 1 plus the assumption that the asymptotic value of R_P is 1.0. This requires that the upper limit of the normal curve integral just mentioned be 2.5 which in turn fixes $\frac{1}{\sqrt{2}\sigma}$ at some value for acquisition, with a new value for extinction and another new value for reacquisition because of shifts in asymptotic values of $(K_C - K_I)HR$.

Figures 1 and 2 indicate that relatively good approximations to trial by trial data can be made by the modified Hullian system suggested. This is not surprising in the case of Fig. 1 since the system of equations used was developed to describe the data therein. Since the theoretical curves of Fig. 2 were devel-

[2] This assumption is *not* equivalent to assuming truncation of the $\dot{E}_C$ and $\dot{E}_I$ distributions and using the resulting $\dot{E}_C - \dot{E}_I$ distribution as Hull might have preferred. The resulting distribution would not be a truncated normal distribution as desired.

oped without any reference to turn behavior except for the number of correct responses made during pretraining, they offer more weighty support to the system.

A noteworthy aspect of Fig. 1 and 2 is that, despite the fact that acquisition levels of speed and correct response were below their asymptotes, extrapolation of the acquisition curves across extinction to reacquisition (just as if rewarded trials had been given instead of extinction) leads to the final reacquisition levels obtained. This phenomenon has been called a "reinforcing effect of extinction" (Cotton, 1957; Jensen & Cotton, 1960) though alternative interpretations are available.

The most striking discrepancies between theoretical and obtained points in these figures occur for early trials of acquisition (running speeds only) and first trials of extinction with both response measures. The early acquisition data are irregular, even though based on 100 *Ss*. Possibly an appeal to frustration drive theory (Amsel, 1958) would permit a better fit of these data. The problem of first-trial extinction data is compensated for by successful prediction of the second-trial increments (during extinction), for speed and for probability of a correct choice, a consequence of *R*'s increase from Trial 1 to Trial 2 of extinction. The discrepancies on Trial 1 of reacquisition suggest that a first-trial parameter varying with spontaneous recovery interval might well be added to the system. Such an addition will not be proposed until more data are available. In view of some indication earlier that only Group 9D differs significantly from the other groups over the 30 trials of reacquisition it is not even certain that parameter changes already made from group to group reflect effects of the independent variable more than of random sampling of *Ss* for the five groups. The principal import of the curve-fitting, then, is to summarize trial-by-trial data in the four stages of the experiment rather than to show well-established effects of the spontaneous recovery interval.

Covariance structure of acquisition speeds.—The existence of reciprocal running times for 100 identically treated *Ss* during acquisition and extinction also permits answering a statistical question relevant to many studies of animal learning. The analysis of variance for repeated measurements first proposed for the two-group case by Lindquist (1947) and then extended for many groups by Kogan (1948) has been proved by Geisser and Greenhouse (1958) to require the assumption of constant variances and constant covariances from trial to trial, though Geisser and Greenhouse do provide a conservative test for use when this assumption is not met. A test of whether this assumption, hereafter called the symmetric covariance assumption, holds is given by Box (1950). Since no evidence in the psychological literature regarding the acceptability of the assumption is known to us, such a test is reported here.

The 30 by 30 variance-covariance matrix for acquisition reciprocal running times[3] was obtained and tested to show whether the symmetric covariance assumption holds. Since a χ^2 of 864.65 with $df = 463$ was obtained, far beyond that required for significance at the .001 level (this χ^2 is equivalent to a z score of 11.17), it seems most unlikely that the assumption holds and hence it seems that the Lindquist-Kogan test is not legitimate for these data. Had such a test been deemed necessary for these data, the conservative procedure of Geisser and Greenhouse would have been necessary.

Fast and slow starting behavior.—Because of informal observations in the Northwestern laboratories of "jerky," "emotional" behavior during extinction, and related findings by Miller and Stevenson (1936), a feature of the present study was the recording of subjective judgments of "fast" or "slow" starts to the runs through the T maze. It was

[3] A 3-page table presenting this matrix and a 1-page table presenting these percentages have been deposited with the American Documentation Institute. Order Document No. 7072 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

thought that extinction runs, though slower in over-all speed, might be marked by fast initial runs followed by refusal to enter the choice point or goal box or by fast retracings to the start box. A comparison of the percentages of fast starts during acquisition, extinction, and reacquisition[3] indicates that this prediction was in error. No statistical analysis was performed on these data, but all group percentages during extinction are lower than any group percentages in acquisition or reacquisition.

SUMMARY

One hundred male albino rats were given identical treatment during 10 pretraining trials with all responses rewarded, 30 acquisition trials with correct responses only rewarded, and 30 extinction trials with no responses rewarded, all in a single-unit T maze. Then randomly assigned groups of 20 *Ss* received spontaneous recovery periods of 1 min., 10 min., 3 hr., 1 day, and 9 days, respectively, prior to a reacquisition session with all responses to the previously correct side rewarded.

Both observed and adjusted (by covariance analysis) mean speeds for the first reacquisition trial indicate systematic increases in running speed from 1 min. through 3 hr., followed by a nonsignificant decline (though not to the 1-min. level) through 9 days. Mean speed increases of 70% or more above the 1-min. condition were shown by the other four groups. Analysis of speeds for all 30 reacquisition trials combined gives similar results except that mean speeds begin to decline after 10 min. and no increment over the 1-min. condition appears in the 1-day or 9-day groups. The corresponding analyses of correct turn data do not reach statistical significance, but a χ^2 test shows that the number of *Ss* having all correct turns following their first correct choice in reacquisition declines significantly with increases in the spontaneous recovery interval.

Neo-Hullian equations are presented and used for fitting mean running speeds and percentage of correct responses for each group on each trial.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

BOX, G. E. P. Analysis of growth and wear curves. *Biometrics*, 1950, **6**, 362–389.

COTTON, J. W. On making predictions from Hull's theory. *Psychol. Rev.*, 1955, **62**, 303–314.

COTTON, J. W. The reinforcing effects of extinction. Paper read at Midwestern Psychological Association, 1957.

COTTON, J. W., LEWIS, D. J., & METZGER, R. Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, **51**, 336–341.

CRESPI L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

DUFORT, R. H., & KIMBLE, G. A. Changes in response strength with changes in the amount of reinforcement. *J. exp. Psychol.*, 1956, **51**, 185–191.

ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.

GEISSER, S., & GREENHOUSE, S. W. An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Ann. math. Statist.*, 1958, **29**, 885–891.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

JENSEN, G. D., & COTTON, J. W. Successive acquisitions and extinctions as related to percentage of reinforcement. *J. exp. Psychol.*, 1960, **60**, 41–49.

KOGAN, L. S. Analysis of variance-repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.

LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend curves. *Psychometrika*, 1947, **12**, 65–78.

MILLER, N. E., & STEVENSON, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205–231.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

YAMAGUCHI, H. G. Drive (*D*) as a function of hours of hunger (*h*). *J. exp. Psychol.*, 1951, **42**, 108–117.

(Received March 30, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 563–567

OPERANT CONDITIONING OF THE GSR [1]

R. L. FOWLER[2] AND H. D. KIMMEL

University of Florida

A controversy regarding the distinction between operant and respondent conditioning (Skinner, 1938) has been whether or not unelicited *autonomic* responses can be strengthened by the presentation of reinforcement after their emission, i.e., as operants. Konorski and Miller (1937) have asserted that autonomic reactions cannot be conditioned as operants, basing their conviction on the grounds that such responses are not instrumental in nature. Skinner (1938), while noting that children learn to cry "real" tears, tended to favor Schlosberg's (1937) notion that autonomic responses are conditioned classically and somatic responses instrumentally.

Kimble (1961) has defended the autonomic-classical and somatic-instrumental dichotomy by citing the unsuccessful attempts by Mowrer (1938) to condition the GSR instrumentally and by Skinner (1938) to condition the vasomotor reflex in operant fashion. Kimble's (1961) conclusion "Thus, for autonomically mediated behavior, the evidence points unequivocally to the conclusion that such responses can be modified by classical, but not instrumental, training methods," (p. 100) is perhaps the most "unequivocal" aspect of the controversy to date.

Recent evidence (Kimmel & Hill, 1960; Mandler, 1960; Razran, 1961) suggests that autonomic responses can be conditioned instrumentally. In the Kimmel and Hill (1960) experiment it was found that the rate of emission of unelicited GSRs was influenced by operant training methods, although the observed differences occurred *after* the reinforcement was terminated rather than during the reinforcement period. In the Russian study reported by Razran (1961) vasodilation (rather than vasoconstriction) became the predominant reaction to electric shock when it was made instrumental in shock termination.

In the present experiment some *S*s received reinforcement following the emission of GSRs while others received the reinforcing stimulus when not emitting GSRs. The reinforcing stimulus was a dim white light presented to *S*s who otherwise received no visual stimulation at all. In addition, the duration of the reinforcement period was varied. It was expected that groups which received reinforcement for twice the length of time would show a greater effect of such reinforcement. It was further expected that the differences during reinforcement, if any occurred, would tend to vanish during extinction.

METHOD

Subjects.—The *S*s were 40 students from introductory psychology classes at the University of Florida. They were assigned to four groups of 10 each, with one restriction on randomness. The first *S* to appear at the laboratory in a given experimental hour was randomly assigned to either a long or a short condition and the next *S* to the opposite duration condition so that 2 *S*s could be run per hour.

Instructions.—The *S*s were instructed to relax and to refrain from abrupt movements

[1] Based upon a Master of Arts thesis done at the University of Florida under the direction of the second author.

[2] Now at the University of Tennessee.

and deep or uneven breathing, but to remain alert and attentive.

Apparatus.—A ¼-in. square light stimulus, produced by a 35-mm. projector located behind a rectangular panel approximately 2 ft. 8 in. in front of *S*'s eyes, was the reinforcing stimulus. Neutral filters were used to attenuate the stimulus. The average illuminance of the light (measured with a Macbeth illuminometer) was 0.36 ft-c and its duration, controlled by a Hunter interval timer, was 1.0 sec.

The GSR was picked up as a dc resistance change from the palm and back of the *S*'s right hand by ¾-in. zinc electrodes, covered with a few drops of saturated zinc sulfate solution, in lucite cups filled with saline electrode paste. It was amplified and then recorded on an Esterline-Angus ink-writing milliammeter with a paper speed of 3 in. per min. The current through *S* was 40 μa. A signal-magnet pen marked the stimulus on the GSR record.

Design.—The independent variables were the contingency or noncontingency of the reinforcement and the duration of the reinforcement period. Contingent lights were presented immediately after *S* emitted a GSR of a predetermined standard magnitude, i.e., at least one-half the average magnitude of that *S*'s unelicited responses during an initial 2-min. rest period. Noncontingent lights were presented only at times when no GSRs were being emitted. Each *S* who received noncontingent lights was matched with a previously run *S* who had received contingent lights, so that the noncontingent *S* received the same number of lights per minute of the reinforcement period as had been received by his contingent pair-mate. Half of the contingent and noncontingent reinforcement *S*s received 8 min. of reinforcement and the other half 16 min. of reinforcement. All *S*s received 4 min. of extinction immediately after the reinforcement period.

Procedure.—When *S* arrived at the laboratory, the palm and back of his right hand were cleaned with alcohol and the electrodes attached. The *S* was then seated in a cushioned chair in an air-conditioned, sound-insulated, electrically-shielded room, and the GSR leads were attached. The instructions were then read and the doors to the experimental room closed, placing *S* in total darkness. The equipment and *E* were located in an adjoining room.

After a brief period during which the GSR amplifier was balanced and its calibration checked, *E* began a 2-min. rest period, marking its beginning and end on the GSR record. The reinforcement criterion for each particular *S*'s response was determined at this time.

The reinforcement period was then begun. The *S*s in the contingent reinforcement groups received the light whenever *E* determined visually from the GSR record that a response met the reinforcement criterion. In making these judgments, *E* tended to be lenient, reinforcing all responses that were close to the criterion. Contingent lights were presented by *E* as quickly as possible after the recorder pen had reached the response magnitude criterion value.

The *S*s receiving noncontingent reinforcement received their stimuli spaced over 1-min. periods according to the response frequency of the contingent group *S* to which each was being matched. Thus, if a particular noncontingent group *S* was to receive four lights during a given minute, the stimuli were arbitrarily spaced out over the 1-min. period at times when no GSRs were being emitted.

After the reinforcement period all *S*s were given a 4-min. extinction period during which no stimuli were presented. Then the equipment was removed and the interview conducted.

Interview.—Each *S* was asked if he had the impression at any time during the experiment that the light was contingent upon "something" he might have done. If a negative reply was given, *E* then told *S* to assume that the light had been controlled by his behavior during the experimental session and, once again, asked if he could think of any possible connection between his behavior and the presence or absence of the light. Only the data of *S*s who completely denied any such correlation were used.[3] The *S*s were then thanked and asked not to discuss the experiment with anyone.

Results

The basic dependent measure was the number of responses emitted per minute of reinforcement and extinction. A response was counted only if its magnitude was approximately equal to or greater than one-half of the average magnitude of the responses emitted by that *S* during the rest period. No response was

[3] Five *S*s, distributed unsystematically among the groups, were replaced for this reason or for failing to follow the instructions regarding movement.

counted if it appeared within 5 sec. after the presentation of a light stimulus (i.e., was elicited by the light). These criteria applied to all *Ss*.

The basic response frequency measure was first transformed to $\sqrt{x+1}$ to overcome the skewness of its frequency distribution. Then, the transformed measures for each minute of reinforcement and extinction were expressed as percentages of the average transformed measure of the 2 min. of rest just preceding reinforcement. This was done separately for each *S*. Figure 1 describes the changes in the measure of transformed relative response frequency for all *Ss* during the period of reinforcement, averaged for 2-min. blocks to reduce variability.

Of primary importance was a comparison of the contingent and noncontingent groups during the reinforcement period. It was expected that the effect of reinforcement would be observable by comparing the relative response frequency curves of the contingent and noncontingent groups. As is indicated in Fig. 1, some separation of the two curves (each based on data of 20 *Ss*) occurred during the first 8 min. of reinforcement, although this difference was not statistically significant ($F = 2.67$, $df = 1/266$) and tended to be obscured by a marked initial tendency in both groups toward reduced response frequency relative to the 2-min. rest period. It was further expected that the divergence of the response frequency curves of the contingent and noncontingent groups would tend to increase as a function of the length of the reinforcement period. Referring again to Fig. 1, it can be seen that even further separation of the two curves (each based on the data of 10 *Ss*) occurred during the last 8 min. of reinforcement.

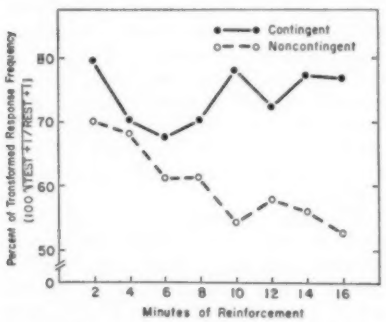

FIG. 1. Percentage of transformed response frequencies during 16 min. of reinforcement ($N = 20$ for each data point during Min. 1–8 and $N = 10$ during Min. 9–16.)

TABLE 1

MEANS AND SDs OF PERCENTAGES OF TRANSFORMED RESPONSE FREQUENCIES [100 $\sqrt{\text{TEST} + 1/\text{REST} + 1}$]

Minutes of Reinforcement	Reinforcement			
	Contingent[a]		Noncontingent[a]	
	Mean	*SD*	Mean	*SD*
Rest[b]	2.44	0.41	2.37	0.44
1	79.0	18.33	73.2	22.34
2	79.6	18.63	65.6	17.57
3	71.0	16.75	63.8	14.52
4	69.8	18.73	71.7	17.87
5	66.3	13.68	63.3	19.51
6	69.0	21.70	59.1	15.52
7	69.9	22.90	61.5	16.18
8	70.8	22.14	62.2	18.18
9	78.6	21.86	56.4	14.64
10	77.6	21.07	52.4	12.66
11	62.2	25.14	58.5	16.93
12	71.8	31.12	57.1	10.28
13	77.9	39.55	59.4	15.29
14	76.5	38.17	53.9	15.14
15	78.8	40.04	55.7	15.48
16	74.6	30.86	50.6	9.65

[a] $N = 20$ for first eight trials and $N = 10$ for second eight trials.

[b] Computed from 2 min. of rest before training, transformed to $\sqrt{X+1}$. These values were used as a baseline for the percentages in the following rows of the table.

Table 1 gives the means and *SD*s of the transformed scores of the contingent and noncontingent groups,

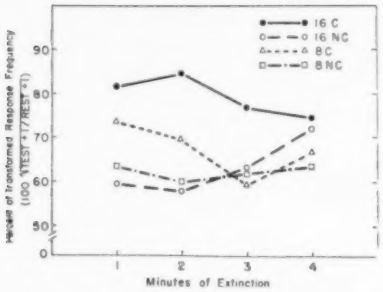

FIG. 2. Percentages of transformed response frequencies during 4 min. of extinction (for each data point, $N = 10$).

for the 16 min. of reinforcement. The analysis of variance for the entire 16 min., using only the measures obtained from the 20 *Ss* who had received 16 min. of reinforcement, showed the Reinforcement Conditions × Minutes interaction to be significant ($F = 1.93$, $df = 15/270$). This interaction reflects the fact that the difference between the two groups grew throughout the reinforcement period, as had been hypothesized.

The data obtained during the extinction period provided another basis for evaluating the hypotheses regarding the effect of reinforcement. Figure 2 depicts the relative response frequency curves for each group separately for the 4-min. extinction period.

It was expected that the two contingent reinforcement group curves would show trends opposite to those of the noncontingent groups during extinction on the assumption that the different reinforcement procedures had opposite effects. Figure 2 suggests, in general, that the contingent and noncontingent curves showed opposite extinction trends, but the analysis of variance of these data provided only marginal grounds for rejecting the null hypothesis of no difference in group trends ($F = 2.25$, $df = 3/108$). The difference between the contingent and noncontingent groups' means was significant only for the first 1 min. of extinction ($t = 2.44$, $P < .025$).

DISCUSSION

The purpose of this experiment was to determine whether or not the GSR could be conditioned as an operant. The data support the conclusion that it can. Both acquisition and extinction response measures showed that the two types of reinforcement, i.e., contingent and noncontingent, led to differences in the rate of emission of unelicited GSRs.

Support was given also to the conclusion that the strength of conditioning depended upon the number of minutes that the unelicited response was reinforced. Finally, it was expected that the trends of the extinction curves for the two contingent subgroups would be opposite to those of the reinforcement curves for the same subgroups. The contingent group which received 8 min. of reinforcement did appear to show a general downward trend during extinction as opposed to an upward trend during the last 2 min. of reinforcement. The contingent group receiving 16 min. of reinforcement showed less systematic change during extinction, probably as a result of having received more reinforcement. That is, the resistance to extinction may have been greater in this subgroup than in the 8-min. contingent reinforcement subgroup. In general, then, the contingent and noncontingent groups' curves diverged during reinforcement and reconverged during extinction, supporting the notion that the two types of reinforcement had opposite effects.

The data presented do conform, in part, to the experimental findings of Kimmel and Hill (1960) and Mandler (1960), and contradict the negative speculations of Schlosberg (1937), Skinner (1938), and the direct assertions of Konorski and Miller (1937), and Kimble (1961).

To eliminate as much as possible,

the possibility that these results were due to "pseudo" operant conditioning (i.e., GSRs mediated by somatic movements which were actually conditioned) careful instructions were given to the *S* before the experiment began and a meticulous interview was conducted with *S* at its termination. All *S*s whose data were used not only stated that they had complied with the instructions, but also rejected the suggestion that their activities may have had any influence upon the occurrence of the reinforcing stimulus.

Summary

This study attempted to determine whether or not the unelicited GSR can be conditioned instrumentally, using a visual reinforcer. The *S*s were 40 college students. All *S*s had 2 min. of initial rest, during which a base response frequency and a response magnitude criterion were determined. A reinforcement period followed the rest period. Two groups received 8 min. or 16 min. of reinforcement following each unelicited GSR. Two additional groups were matched to the first two groups in the number of visual stimuli presented per minute of reinforcement, but the stimuli were not given contingent upon the occurrence of unelicited GSRs. Finally, a 4-min. period of extinction was given. A significant interaction between reinforcement conditions and minutes of reinforcement was found in the 16-min. groups but not in the 8-min. groups. In general, the groups receiving contingent lights tended to respond at a higher rate than the noncontingent groups. During extinction, the relative response frequency trends for the contingent groups' curves tended to be opposite to those of the noncontingent groups (downward and upward, respectively), although this effect was not statistically significant. The results were interpreted as supporting the conclusion that the unelicited GSR can be conditioned instrumentally.

References

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

Kimmel, H. D., & Hill, F. A. Operant conditioning of the GSR. *Psychol. Rep.*, 1960, **7,** 555–562.

Konorski, J., & Miller, S. On two types of conditioned reflex. *J. gen. Psychol.*, 1937, **16,** 264–272.

Mandler, G. *Final report on Grant M-2442.* Washington: United States Public Health Service, 1960.

Mowrer, O. H. Preparatory set (expectancy): A determinant in motivation and learning. *Psychol. Rev.*, 1938, **45,** 62–91.

Razran, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, **68,** 81–147.

Schlosberg, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, **44,** 379–394.

Skinner, B. F. *The behavior of organisms: An experimental analysis.* New York: Appleton-Century, 1938.

(Received April 15, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 568–574

AN ANALYSIS OF SET IN RELATION TO REACTION TIME

JACK BOTWINICK AND JOSEPH F. BRINLEY

National Institute of Mental Health

Statements or inferences about set (expectancy or readiness) are made from the relation between reaction time (RT) and the foreperiod or preparatory interval (PI) in a series. For example, Woodworth and Schlosberg (1954), in referring to an early study by Woodrow (1914), reported that "Maximum readiness . . . was not reached in much under 2 seconds and was not maintained much longer than 2 additional seconds" (p. 30). The reference was to RTs of 3 *S*s with a regular series.

In this statement by Woodworth and Schlosberg, set was described by two aspects—reaching and maintaining. RT set has also been described in terms of developing, organizing, mobilizing, and other aspects (e.g., Botwinick, Brinley, & Birren, 1957; Botwinick, Brinley, & Robbin, 1959; Huston, Shakow, & Riggs, 1937; Rosenthal, Lawlor, Zahn, & Shakow, 1960). In the present study the hypothesis that the various aspects of set are distinguishable, different in meaning, and independent was tested by principal component analyses (Hotelling, 1933). The analyses were based upon the intercorrelations of RTs for various PIs within various contexts. For example, if a "reaching" of set, as inferred from the RTs to short PIs, is independent from a "maintaining" of set, as inferred from RTs to long PIs, then clear components loaded by different PI durations might be expected. One component might be of short or relatively short PIs, and another component of long PIs.

Statements or inferences about set are also made from relations among RT-PI functions. For example, Huston, Shakow, and Riggs (1937), and Botwinick, Brinley, and Robbin (1959), compared RTs of regular and irregular series and discussed the relative slopes of the functions in terms of maintaining set. It is known that the value of RT in irregular series is related to the context in which the PI is embedded (e.g., Botwinick et al., 1957; Karlin, 1959; Klemmer, 1956), although its absolute duration may be a limiting factor (Botwinick, 1959, p. 749). Context even seems to make a difference in regular series, at least for schizophrenic *S*s, since PIs given in ascending order (increasing PI lengths) made for RTs that were different from those of PIs given in descending order (Zahn, Rosenthal, & Shakow, 1961). Thus the relations among parameters of RT-PI functions would appear to be contingent upon the specific contexts of the PI, as well as the absolute duration of the PI itself. In the present study, an attempt was made to assess the relative contribution of the absolute PI duration to that of its context. The same PIs in different series, varied in range and central tendency, were compared with respect to the associated RTs.

A third aspect regarding RT set involves sensory modality. While the particular sense modality is a factor in overall level of RT, it is not thought to be a determinant of set itself. Thus aspects of set are assumed to be independent of the sensory time courses. This assumption was tested as a hypothesis by com-

paring visual and auditory RTs in relation to the PIs of two types of series. In addition, these RTs were analyzed in relation to visual and auditory acuity of *S*.

METHOD

Subjects.—The *S*s were 32 men and 32 women volunteers. Mean age was 20.9 yr. (range 18 to 27 yr.), and mean education was 13.7 yr. (range 12 to 17 yr.). Twenty-seven of these *S*s (25 men and 2 women) were university students who came to the laboratory and were paid for their participation. The remainder of *S*s was quartered in an NIH ward for research purposes. All *S*s were in apparent good health with no known disease.

Procedure.—RT measurements were made with each *S* during three different sessions, each session lasting about 40 min. and taking place in most instances within 1 week, but never taking longer than a 2-week interval.

The total three-session period included 300 RT measurements under four conditions of auditory stimulation, and two conditions of visual stimulation. The four auditory conditions comprised three irregular series and a regular series. One regular and one irregular series comprised the visual series.

The visual and auditory regular series had PIs of 0.5, 1.0, 3.0, 6.0, and 15.0 sec. Half the *S*s were given the series in ascending order, and half in descending order. One auditory irregular series and one visual irregular series comprised these same five PI durations.

The remaining two auditory irregular series were of the following PIs: 0.5, 1.0, 1.5, 2.0, and 3.0 sec., and 3.0, 6.0, 9.0, 12.0, and 15.0 sec. All irregular series were arranged so that each PI followed every other PI an equal number of times.

All five PIs within each of the six series were presented 10 times. In the irregular series, in order to have each PI follow every other PI the same number of times, two additional measurements were made with the first and the twenty-seventh RT measurement (of the 52 in the series) not counted in the analysis.

Brief practice or warm-up RTs were given at the start of each series. Irregular series were preceded by two presentations of each of the five PIs in random order. Each PI of the regular series was presented twice with the 10 RT measurements for the PI commencing immediately after this.

Each of the three sessions comprised two series: one session was of the two visual series, another session was of the two comparable auditory series, and one session was of the auditory irregular series where the PIs ranged from 0.5 to 3 sec., and 3.0 to 15.0 sec. The order of the three sessions was systematically varied for the 64 *S*s, and the two series within the three sessions were counterbalanced as nearly as possible.

The *S* pressed down on a telegraph response key which started an automatically programed sequence. The finger press initiated a warning signal of 0.5 sec. duration, followed by the PI and then the stimulus. The stimulus was an 85-db., 1000-cycle tone presented through earphones, or the glow of a neon 40-w. lamp in front of *S*. The stimulus was terminated by the finger lift response. The warning signal was the glow of a 7½-w. frosted lamp in the case of the auditory stimulus, or a buzzer, approximately 64 db., in the case of the visual stimulus. Each *S* was encouraged to respond as quickly as possible.

Measures of auditory and visual acuity were made on each *S*. Air conduction audiograms were obtained with a Western Electric Audivox 7-B audiometer. Hearing loss in decibels of a 1000-cycle tone was measured for each ear in three ascending and three descending series. The median of the six series was used as the measure of hearing loss. Visual acuity measures were made of near and far vision (optical distances of 13 in. and 26 ft., respectively) with a Bausch and Lomb Orthorater. Standard test slides of binocular acuity were used.

RESULTS

Ten RT measurements were made for each PI within each series for each *S*. The median of each of these 10 measurements was computed and used in the analyses of the data.

PI in context.—It is clear from Fig. 1 that RT was related to absolute PI duration within a series, and to PI duration relative to the specific PI context. Thus for example, the four RTs with a PI of 3 sec. ranged in value from .172 to .213 sec. The four mean RTs were significantly different ($P < .01$) as determined by a variance analysis, and each of the

four mean RTs were significantly different from each of the other three ($P < .02$–$.01$) as determined by t tests. In general, regular PIs made for quicker RTs than did irregular PIs although there was no reliable RT difference between the regular and irregular series with PIs of 15.0 sec. ($P > .05$). Reliable differences between RTs of regular and irregular series were found with the other four PIs of the series ($P < .01$).

The data in Fig. 1 also showed that statements regarding specific aspects of set based upon the slope of the functions need to be made in reference to the particular PIs in their context. Thus, the magnitude of the difference in RT between PIs of 1.0 and 0.5 sec. was contingent upon the range or central tendency of the series. Similarly, differences in RT between PIs of 6.0 and 3.0 sec. were contingent upon context. In each of these cases statistically significant differences were found between each RT difference tested against the other two ($P < .05$–$.01$). However, the absolute duration of PI, independent of context, appeared to be a limiting factor, at least with long intervals. With abscissa values longer than 6.0 sec., the top two curves in Fig. 1 followed very similar courses, and with PIs of 15.0 sec., the three RT values were almost identical.

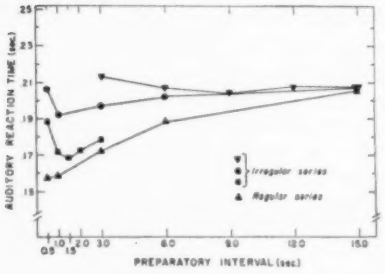

FIG. 1. Mean simple auditory RT of 64 *Ss* as a function of the preparatory interval in four different series contexts.

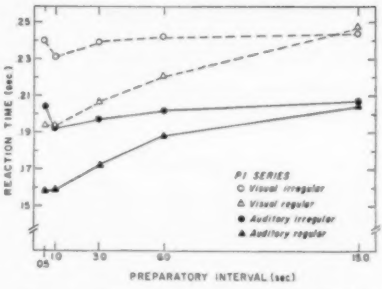

FIG. 2. Mean simple auditory and visual RT of 64 *Ss* as a function of five preparatory intervals in regular and irregular series.

Sense modality.—It may be seen in Fig. 2 that RT was longer with the visual stimulus than with the auditory stimulus but that the increased latency was not a function of PI or its context. The data of auditory RT in this figure are the same as the corresponding data in Fig. 1. An analysis of variance was performed with the data of Fig. 2, but men and women *Ss* were kept unpooled.[1] As shown in Table 1, the interactions between sense modality and PI, sense modality and type of series (i.e., regular and irregular), and sense modality, PI, and type of series, were not statistically significant ($P > .05$). However, visual RTs were longer than auditory RTs ($P < .01$).

In addition to the main effect of sense modality, the variance analysis included the main effects of PI, type of series, and also sex of *S*. It may be seen in Table 1 that all main effects and only one interaction (PI×Series) were significant ($P < .01$). This

[1] We thank Seymour Geisser, Section on Mathematics and Theoretical Statistics, National Institute of Mental Health, for directing the variance analysis, and we thank Annie T. Randall, of the same section, for the computation involved.

significant interaction indicated what may be seen in Fig. 2, viz., that the nature of the relation between RT and PI is different for the two types of series. Women *Ss* were generally slower than men *Ss*, different PIs made for different RTs, and the regular series made for generally quicker RTs than did the irregular series.

In the next section the overall results of principal component analyses will be described. One of these analyses comprised the data of Fig. 2. As with the variance analysis, sense modality appeared independent of the PI. Separate components of visual and of auditory RT were not found.

Correlations of the visual and of the auditory acuity measures with the respective RTs ranged from low to zero. Visual and auditory acuity measures were also uncorrelated.

TABLE 1

ANALYSIS OF VARIANCE OF RTs TO VISUAL AND AUDITORY STIMULI

Source	*df*	*MS* (10000)	*F*
Sense mode (SM)	1	4545	544.98*
Prep. interval (PI)	4	420	98.73*
Series type (ST)	1	2020	185.50*
Sex (S)	1	274	21.21*
Individuals (I) within S	62	13	
SM × PI	4	1	0.55
SM × ST	1	8	1.90
SM × S	1	0	0.00
PI × ST	4	194	27.67*
PI × S	4	5	1.15
ST × S	1	1	0.11
SM × I	62	8	
PI × I	248	4	
ST × I	62	11	
SM × PI × ST	4	3	0.30
SM × PI × S	4	3	1.62
SM × ST × S	1	1	0.13
PI × ST × S	4	7	1.00
SM × PI × I	248	2	
SM × ST × I	62	4	
PI × ST × I	248	7	
SM × PI × ST × S	4	1	0.11
SM × PI × ST × I	248	11	
Total	1279		

* $P < .01$. All other F ratios not significant at .05 level.

TABLE 2

PRINCIPAL COMPONENTS I AND II OF SIX PREPARATORY INTERVAL (PI) SERIES: UNROTATED MATRIX

PI (Sec.)	Visual RT		Auditory Reaction Time			
	Reg.	Irreg.	Reg.	Irreg.	Irreg.	Irreg.
	Components I					
0.5	65	81	53	87	89	
1.0	82	86	80	88	91	
1.5					89	
2.0					88	
3.0	85	88	91	87	91	89
6.0	90	85	85	85		93
9.0						93
12.0						90
15.0	84	89	80	88		94
%[a]	67	74	63	76	81	84
	Components II					
0.5	67	48	78	32	39	
1.0	28	28	30	35	18	
1.5					−32	
2.0					−09	
3.0	−06	−23	−06	−16	−17	39
6.0	−31	−39	−30	−22		18
9.0						−18
12.0						−35
15.0	−40	−11	−42	−29		−04
%[a]	16	11	19	08	06	07

Note.—Loadings have been rounded from six places and decimals omitted.
[a] Percentage variance explained by component.

Components of set.—A Hotelling (1933) principal component analysis was performed on the RTs of the five PIs in each of the six PI series.[2] In each of the six component analyses, a first component was found which accounted for a large percentage of the RT variance.[3] In Table 2, it may be seen that

[2] The intercorrelations and the principal component analyses were directed by Donald F. Morrison of the Section on Mathematics and Theoretical Statistics, National Institute of Mental Health. IBM 650 digital computers were used for both the intercorrelations and the component analyses. We are grateful to Donald Morrison for this kind service.

[3] Copies of the correlation and principal component matrices will be furnished by the National Institute of Mental Health on request.

the variance associated with these first components ranged from 63% to 84%, and all PIs within a series loaded positively and highly. The loadings within each of the four irregular series were homogeneous with differences between them being only .03 and .05 in the auditory series, and .08 in the visual series. The loadings on the first component of the regular series were also homogeneous except for the shortest PIs (0.5 sec.). While the loadings of the 0.5-sec., regular PIs were high, they were substantially lower than those of the other four PIs in the series.

In Table 2 the second component for each of the six PI series may be seen. The variance accounted for by these components ranged from 6% to 11% in the four irregular series, and was 16% and 19% for the visual regular and auditory regular series, respectively. In each of the six series, the second component appeared to be related to PI duration, having high positive loadings for the shortest intervals and negative loadings for the longest intervals. Subsequent components, not included in Table 2, were uninterpretable, being less systematic in their bipolarity than the second components.

The polarity associated with PI length in the second components suggested that with orthogonal rotations of axes, components could be obtained which reflected independent variance of the shortest PIs. Table 3 presents the orthogonal solutions by graphic rotations of the attempt to obtain such components. The rotated loadings involved the first two or three principal components of each of the six series.

The first component in each series was once again of general RT or RT set. The percentage of variance accounted for by each of these first components ranged from 58% to 81% with loadings of individual PI durations being relatively homogeneous, except again, for the 0.5-sec., regular PIs.

The second components of the six series were clearly identified with the short preparatory intervals. The rotations resulted in high loadings of short duration PIs and near zero loadings of longer ones. This was the case independent of regularity or irregularity of interval, and of the range and magnitude of PI durations within the series. Thus, both the absolute and relative shortness of the interval within a series was important in the organization of the short interval components, although the loadings of the shortest intervals were appreciably highest in the two regular series. The 0.5-sec., regular PIs of the short interval components accounted for more of the communality than these same PIs of general components. In all six series, however, the second components indicated that preparation during short intervals was distinguishable and independent of general RT.

In an attempt to discover short interval components common to more than one PI series, a principal component analysis was performed on RTs of the two regular series together.[3] Again, after rotation, a second component was found which was clearly identified as one of short interval. There were high loadings for the shortest PIs of both the visual and the auditory series, and negligible loadings for the longer intervals of both series. There was evidence of a similar nature for a short interval component common to the three irregular series which included the 0.5 sec. PIs.[3]

TABLE 3

PRINCIPAL COMPONENTS I AND II OF SIX PREPARATORY INTERVAL (PI) SERIES: ROTATED MATRIX

PI (Sec.)	Visual RT		Auditory Reaction Time			
	Reg.	Irreg.	Reg.	Irreg.	Irreg.	Irreg.
Components I						
0.5	44	66	27	76	78	
1.0	77	76	75	76	85	
1.5					94	
2.0					88	
3.0	87	91	90	88	92	78
6.0	93	92	86	88		88
9.0						94
12.0						95
15.0	87	89	86	92		93
%[a]	63	70	58	71	77	81
Components II						
0.5	89	67	97	53	58	
1.0	30	49	28	56	38	
1.5					−11	
2.0					11	
3.0	02	00	16	06	04	55
6.0	00	−16	08	00		35
9.0						00
12.0						−17
15.0	−01	12	−09	−06		14
%[a]	18	15	21	12	10	10

Note.—Decimals have been omitted from the loadings.

[a] Percentage variance explained by component.

When both regular and irregular series were considered together, evidence did not appear for components loading on the shortest intervals. The RTs of all 30 PI conditions were intercorrelated and subjected to a principal component analysis. In addition, a component analysis was made of a 34 × 34 matrix in which the two visual and two auditory acuity measures were included. Component analyses were made also of RTs of the 20 PI conditions in Fig. 2 including the four acuity measures, and of the auditory RTs in Fig. 1 including the two auditory acuity measures.[3]

The first components of these four analyses were similar to each other, and to the first components of the six individual series. The first component in each of the four analyses accounted for approximately half the total variance, and all PIs loaded heavily on these components. Again, the loadings of the 0.5-sec., regular PIs were lower than the others. While the first components were easily identifiable, viz., general, the subsequent components were uninterpretable in their complexity. There was bipolarity that had no apparent order, there was no evidence of separate visual or auditory loadings, but there was some indication of regularity and irregularity of PI condition. Unlike the analyses of individual series, attempts to rotate the axes did not clarify the structure.

The 34 × 34 matrix of correlations indicated relations that were not apparent from the principal component analyses. Within the four irregular series, the RTs were intercorrelated highly among the PIs and were relatively homogeneous in magnitude. Within the two regular series, however, there were both high and low correlations, ordered with respect to the duration of PI. Thus, in the auditory regular series, the coefficient of correlation was .73 between RTs to the 15.0- and 6.0-sec. intervals. The coefficient was only .17 between 15.0- and 0.5-sec. intervals. The corresponding correlations among the visual regular RTs were .85 and .36, respectively.

Discussion

The principal component analyses were performed to test the hypothesis that various aspects of RT set were distinguishable, different in meaning, and independent. In whatever PI series or group of PI series analyzed, a first component was found which accounted for an extensive amount of variance and which was clearly identifiable as a general RT or RT set component. These first components did not disappear on rotation. Short-interval components were also found, appearing as the second components in the rotated matrices of individual PI series, and in some of the series analyzed together. The short-interval components suggested at least one distinguishable aspect of RT set independent of overall set or RT level. The number of distinguishable aspects of RT set that were different in meaning, however, appeared to be limited in the present data. What other aspects were present, were ambiguous and too complex for identification. Long-interval components were not found, nor were components of meaningful combinations of long and short PI durations.

The meaning of the short-interval components appeared similar or general within regular series and within irregular series. In the two regular series together, and in the combination of irregular series involving the 0.5-sec. PIs, the shortest intervals within a series were grouped together. Thus, in regular series, the short interval components emphasized ability to become set for quick response within short time durations, and this ability was independent of the one for quick response when the interval indexing set was longer. The specific duration of the PI for this latter ability was of relatively little consequence. In irregular series, the short-interval components suggested ability to become set for quick response within PI durations that were short *relative* to other PIs in the series.

The short-interval components of the two regular series were larger in the proportion of variance explained than were those of the four irregular series, and perhaps were more important because of the very high loadings of the 0.5-sec., regular PIs. The meaning of

the short interval may be different for the two types of series. Figures 1 and 2 indicate that RT was minimal with the 0.5-sec. PI in regular series but not in irregular series. When regular and irregular series were analyzed in one matrix, the 0.5-sec. PIs of the different series did not contribute to the variance of the components in the same way. In both types of series, however, RT variance associated with PI duration was found independent of overall RT variance.

The second aspect of RT set in this study involved the context of the PI. There is little doubt that RT was related to both the PI and its specific context, although context may be less important with long PI durations than with short ones. Thus, aspects of set, inferred from slope and other parameters of RT-PI functions, are related to PI contexts, and statements of RT set regarding populations, or comparisons of populations, must be made in reference to the specific conditions of measurement.

The remaining aspect of RT set investigated involved the role of sensation. Sense modality was a factor in overall level of RT, but not related to set as examined in this study.

Summary

Principal component analyses were performed with reaction times (RTs) in relation to preparatory intervals (PIs) in six different series. Components of general RT level or RT set were found on which all PIs loaded substantially. In addition, short-interval components were found when matrices were rotated orthogonally. Long-interval components were not found, nor were other identifiable components. The results were discussed in relation to the hypothesis that aspects of RT set were distinguishable, different in meaning, and independent.

RTs were related to both the PI and its specific context. While RT level was different for auditory and visual stimulation, the difference was not related to PI or its context. Thus, set, as inferred from the relation between RT and PI, was independent of sense modality.

References

Botwinick, J. Drives, expectancies, and emotions. In J. E. Birren (Ed.), *Handbook of aging and the individual: Psychological and biological aspects.* Chicago: Univer. Chicago Press, 1959. Pp. 739–768.

Botwinick, J., Brinley, J. F., & Birren, J. E. Set in relation to age. *J. Gerontol.*, 1957, **12**, 300–305.

Botwinick, J., Brinley, J. F., & Robbin, J. S. Maintaining set in relation to motivation and age. *Amer. J. Psychol.*, 1959, **72**, 585–588.

Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, **24**, 417–441, 498–520.

Huston, P. E., Shakow, D., & Riggs, L. A. Studies of motor function in schizophrenia: II. Reaction time. *J. gen. Psychol.*, 1937, **16**, 39–82.

Karlin, L. Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, 1959, **58**, 185–191.

Klemmer, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

Rosenthal, D., Lawlor, W. G., Zahn, T. P., & Shakow, D. The relationship of some aspects of mental set to degree of schizophrenic disorganization. *J. Pers.*, 1960, **28**, 26–38.

Woodrow, H. The measurement of attention. *Psychol. Monogr.*, 1914, **17**(5, Whole No. 76).

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1954.

Zahn, T. P., Rosenthal, D., & Shakow, D. Reaction time in schizophrenic and normal subjects in relation to the sequence of series of regular preparatory intervals. *J. abnorm. soc. Psychol.*, 1961, **63**, 161–168.

(Received April 21, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 575–578

DONDERS' B- AND C-REACTIONS AND S-R COMPATIBILITY [1]

D. E. BROADBENT AND MARGARET GREGORY

Applied Psychology Research Unit, Cambridge, England

It was held by Donders (1868) that reaction time (RT) under b-conditions is longer than reaction time under c-conditions. That is, if a man is presented with two stimuli repeatedly in a random order, and if he has to make one reaction to one stimulus and a different reaction to the other stimulus, his RT is longer than it would have been if his instructions were to react only to one stimulus and to remain inactive to the other. This result has sometimes been regarded as demonstrating that the choice between two actions represents an extra process in addition to that of discriminating two stimuli. Donders himself suggested that the time taken by this extra process could be measured by b-c; but this involves the assumption that the two processes do not overlap in time, and this has usually been felt to be doubtful. The basic observation, however, is not normally questioned.

During the course of an investigation into verbal RTs, it was noticed that two groups of *S*s gave the same average time despite the fact that one was performing a b-reaction and one a c-reaction. Since the groups were small, this equality might have been due to chance factors; alternatively it was possible either that the classic experiments on the topic contain some loophole, or else that verbal RTs were abnormal in this respect. The classic experiments are not indeed very convincing to modern eyes. If for instance more than two stimuli are used, each receiving a separate response in the b-reaction but all except one ignored in the c-reaction, then there is a reduction in the number of stimuli discriminated as well as an abolition of choice between reactions. It is now of course realized (Crossman, 1953; Hick, 1952; Hyman, 1953) that such a reduction in the information transmitted through the nervous system may allow a reduction in RT quite apart from any other factor. However, there are cases in the literature in which only two stimuli were used and yet the usual difference between b- and c-reactions was found. Hick (1952), for example, in giving data for varying numbers of alternative reactions, includes the case where only one reaction is made. It appears from the experimental procedure that this case was in fact a c-reaction using the same two stimuli as the two-choice reaction, which is therefore the corresponding b-condition. The difference amounts to as much as 100 msec. in favor of c-reaction, so that it is quite clear and unambiguous. Thus the classic statements appear to be true even though the evidence which gave rise to them is not always convincing in the light of later knowledge.

There remains the possibility that verbal reactions are unusual. It may be recalled that Fitts and Seeger (1953) established that RT was faster if the type of response to be made was one which was particularly appropriate to the stimulus presented. That is, if a set of stimulus lights are arranged in a square, it is best for the response keys also to be arranged in a square, with each key corresponding to the light in the same position. The extent to which the responses were related to their stimuli in such a direct way was termed by Fitts and Seeger (1953) the "compatibility" of the S-R relationship. Now some recent evidence has suggested that an increase in the compatibility of a situation may abolish the usual tendency for RT to increase as the number of alternative reactions increases. Leonard (1959), for instance, has found that in responding to stimulation of a finger by pressing with that same finger, RT is largely independent of the number of fingers that must be discriminated. Pierce and Karlin (1957) found that the rate of reading, of English words seen clearly in

[1] Thanks are due to the British Medical Research Council for supporting this research; and to J. A. Leonard, not only for theoretical discussions, but for the loan and modification of an apparatus designed by him.

good illumination, was not affected by the size of the vocabulary in use, and it can perhaps be accepted that the most compatible response to a printed word is to say that same word. Mowbray and Rhoades (1959) have demonstrated for 1 *S* that, with very high degrees of practice, the difference between the time to choose one of four alternatives, and the time to choose one of two, will disappear even when the situation is the usual laboratory one of pressing keys in response to light signals.

Finally, Davis, Moray, and Treisman (1961) have shown that the time taken to repeat a word which is heard (a highly compatible task) does not depend upon the size of the vocabulary in use, and the unduly high speed of multiple choice reactions performed in this way has been observed independently by Chistovich, Aliakrinsky, and Abulian (1960).

While these workers have thus been indicating that only in relatively incompatible situations does RT increase with the number of alternative reactions, Rabbitt (1959) has shown that in such an incompatible situation it is the number of responses rather than the number of stimuli which primarily affects RT. If the same response is to be made to each of several different stimuli, it makes little difference how many stimuli are in fact involved so long as the number of responses is constant. Thus it is tempting to consider that the difference between the b- and c-reactions is of the same kind as that between two- and four-choice situations, both differences being due to the need to keep varying numbers of response tendencies in readiness. If this were so, the b- and c-conditions might well be identical in a highly compatible situation such as repeating heard words.

The present experiments were therefore designed to investigate the following questions: (*a*) Can the b- and c-reactions indeed be equally fast when the situation is one of repeating heard words? (*b*) Are the two conditions also equally fast in another highly compatible situation not involving the speech mechanism? (In fact, Leonard's tactile stimulation was employed.) (*c*) For both types of stimulus, does the traditional difference between b- and c-reactions reappear if the situation is made incompatible by changing the S-R relationship?

Method

Auditory reactions.—Stimuli were presented from a Ferrograph tape recorder through a loudspeaker mounted in a bass-reflex cabinet facing and touching *S* at knee height. The playback level was a comfortable listening one, i.e., it produced peak readings of about 75 db. on a Dawe sound-level meter. The responses were picked up by a Reslo ribbon microphone placed on top of the speaker cabinet and thus close to *S*'s mouth. Both stimulus and response were recorded on a second Ferrograph, and the time interval between them could afterwards be measured by locating them auditorily using small manual movements of the tape past the playback head: and measuring distance along the tape. This method of measurement possesses the advantage that the person making the measurement can ensure that the part of the response used is an informative one; and not merely a meaningless preliminary to speech as it may be with a voice key. On the other hand it is of course restricted to giving individual RTs to the nearest 10 msec. (about 2 mm. of tape).

The stimulus tape consisted of the words BID and DID arranged in random order, subject to the restriction that equal numbers of both stimuli occurred in each block of 30, and that no runs of more than 3 identical stimuli occurred. The interval between successive stimuli was 5 sec., and each was preceded by the word READY approximately 2 sec. earlier. There were four blocks, so that the whole tape totaled 120 stimuli, and the same tape was used for all conditions.

Tactile reactions.—The apparatus has previously been described by Leonard (1959). Briefly, *S* rested his fingers on relays, and a stimulus was given by applying 50-cps ac to the relay to produce a vibration. The response was pressure downwards on the relay. In the present experiment only the two index fingers were used, and the two corresponding relays were fitted each with a metal ring into which the finger was inserted. This prevented *S* from lifting one finger off the relay in the c-condition and so avoiding the stimulation of that finger altogether. RT was measured to the nearest 10 msec. by a Dekatron timer.

The sequence of stimuli used was identical to that for the verbal reactions, replacing BID by stimulation of one hand and DID by stimulation of the other.

Procedure.—Four groups of *S*s were tested,

TABLE 1

MEAN REACTION TIME (IN MSEC.) UNDER EACH CONDITION

Response	S-R Compatibility	*N*	b-Cond.		c-Cond.		*Ss* Giving b > c
			Mean	*SD*	Mean	*SD*	
Verbal	Compatible	11	284	55.4	292	31.1	4
	Incompatible	6	444	31.2	354	57.7	6*
Tactile	Compatible	12	193	24.2	194	28.1	6
	Incompatible	6	300	25.3	255	19.5	6*

* $P < .05$ for b − c.

all Royal Navy enlisted men. No *S* served in more than one group. Each group was tested on 2 successive days, half the *Ss* receiving the b-condition on Day 1 and the c-condition on Day 2, and half reversing this order. On each day a run of 120 stimuli was given, with a rest between each block of 30. The first 10 responses to the word BID (or with the right finger, as appropriate) were scored in each of the last two blocks. No difference in the trend of the results appeared between these blocks, and they have therefore been combined.

Group VC (Verbal Compatible) ($N = 11$) was instructed in the b-condition to respond to either word by saying the same word as was heard. In the c-condition they were to do this only for the word BID and were to ignore the word DID.

Group VI (Verbal Incompatible) ($N = 6$) was instructed in the b-condition to respond to the word BID by saying the word DECK, and to respond to the word DID by saying the word DIM. In the c-condition the word BID only was to receive the response DECK and the other stimulus was to be ignored.

Group TC (Tactile Compatible) ($N = 12$) was instructed in the b-condition to press down with the same index finger that was stimulated. In the c-condition they were to ignore stimulation of the left index finger and to respond only to the right finger.

Group TI (Tactile Incompatible) ($N = 6$) was instructed in the b-condition to press down the left index finger in response to stimulation of the right hand, and vice versa. In the c-condition they were to ignore stimulation of the right hand, and merely to respond with that hand to stimulation on the left side.

RESULTS AND DISCUSSION

Table 1 gives the mean RTs for each group in each condition. It is clear that both the groups working under compatible conditions give b-reactions that are slightly faster than their c-reactions. The difference is completely insignificant ($P > 0.5$ by a nonparametric test based on tau), and is in any event in the opposite direction to the traditional finding. The groups under incompatible conditions however give much faster c-reactions than they do b-reactions, and in each group every *S* shows this difference so that for each group $P < .05$ by sign test. Thus the traditional result seems to appear only with rather incompatible S-R relationships.

The RTs obtained are comparable with those found by Leonard (1959) in comparable conditions.

These results form part of the general body of data, cited in the introduction, which show that S-R compatibility interacts with the degree of choice present in a reaction time task. The theoretical implications of this empirical generalization have been discussed by Leonard (1960). It suggests a theoretical model of the type discussed mathematically by Stone (1960). Such models may be described as regarding the tendency to make any particular reaction as varying from moment to moment about a mean value, the variance reducing with the time for which the reaction is delayed. If the mean value is greater for the correct reaction than for erroneous ones, and if the reaction actually occur-

ring is the one with the strongest tendency at the time chosen for response, RT will (for a constant probability of error) increase both with the number of erroneous reactions which are possible in the situation, and with the strength of the tendency to make any particular error. Such models include, as a special case, the familiar proportionality of RT to stimulus information: but they are of course related to decision theory rather than to the traditional informational analysis. In the present case, they would predict that the effect of adding an extra response to a task will depend upon the extent to which that response is one which competes strongly with the response already required in the task. The results are therefore consistent with a model of the decision theory type.

Summary

It was hypothesized that the increase in RT under b- as opposed to c-conditions would vary according to the S-R compatibility of the particular situation used. Four groups of *Ss* were therefore tested, two using verbal reactions to verbal stimuli, and two using key reactions to tactual stimulation of the fingers. One of the groups having each type of response used a compatible S-R relationship, i.e., they responded with a word identical to the word heard, or pressed a key using the actual finger stimulated. These groups showed no difference between the b- and c-conditions of Donders (1868). The other two groups used incompatible relationships, responding with a word other than the one heard, or with a finger other than the one stimulated. These groups showed the classic difference between b- and c-conditions. It can therefore be concluded that this difference is greatly reduced in conditions of high S-R compatibility.

References

Chistovich, A. A., Aliakrinsky, V. V., & Abulian, B. A. Time delays in speech repetition. *Prob. Psychol.*, 1960, **1**, 64–70.

Crossman, E. R. F. W. Entropy and choice-time: The effect of frequency unbalance on choice response. *Quart. J. exp. Psychol.*, 1953, **5**, 41–51.

Davis, R., Moray, N., & Treisman, A. Imitative responses and the rate of gain of information. *Quart. J. exp. Psychol.*, 1961, **13**, 78–89.

Donders, F. C. Die Schnelligkeit psychischer Processe. *Arch. Anat. Physiol., Leipz.*, 1868, 657–681.

Fitts, P. M., & Seeger, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.

Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Leonard, J. A. Tactual choice reactions. Part 1. *Quart. J. exp. Psychol.*, 1959, **11**, 76–83.

Leonard, J. A. Choice reaction time experiments and information theory. *Proc. Lond. Symp. Inform. Theory*, 1960, **4**.

Mowbray, G. A., & Rhoades, M. V. On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11**, 16–23.

Pierce, J. R., & Karlin, J. E. Reading rates and the information rate of a human channel. *Bell Sys. tech. J.*, 1957, **36**, 497–516.

Rabbitt, P. M. A. Effects of independent variations in stimulus and response probability. *Nature*, 1959, **183**, 1212.

Stone, M. Models for reaction time. *Psychometrika*, 1960, **25**, 251–60.

(Received April 24, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 579–588

TASTE AND CONSUMMATORY ACTIVITY IN AMOUNT AND GRADIENT OF REINFORCEMENT FUNCTIONS[1]

FREDERICK A. KNARR AND GEORGE COLLIER

University of Missouri

In a prior runway study using sucrose reinforcements (Collier, Knarr, & Marx, 1961) it was concluded that variations in the total amount of consummatory activity are not necessary conditions for reinforcement. The present study is concerned with the question as to whether such variations are sufficient conditions. The taste of the reinforcing substance is varied conjointly with the amount of consummatory activity required to obtain the reinforcement.

One implication of the hypothesis that the behavior in the runway is, in part, chained by surrogates of the consummatory response is that the structure of the chain will be some function of the magnitude of the consummatory response. The present study examines this implication in terms of the variation in the gradient of running as a function of variations in the amount of consummatory activity as well as taste.

METHOD

Experiment 1

Subjects.—The *Ss* were 53 male and 48 female albino rats, 90–115 days old at the start of the study. Twelve groups were matched on the basis of sex and predeprivation weights. One female from each group and a total of 5 males were discarded. Of the discarded *Ss*, 3 *Ss* failed to learn, 2 died, 2 were lost by procedural errors, and 10 were randomly discarded to equalize the number (7) of *Ss* per group.

Apparatus.—A straight runway (Collier et al., 1961), 18 ft. long, 3¼ in. wide, and 5 in. high, enclosed in a light-tight, sound-deadened hull was used. The start box occupied the first 1½ ft., the stem the following 14 ft., and the end box the remaining 2½ ft. Guillotine-type doors, operated by motor driven screws, separated the start box and the end box from the stem. The floor consisted of nine pairs of 2 ft. long, 1½ in. wide pieces of aluminum angle separated by $\frac{3}{16}$ in. side to side and by $\frac{1}{16}$ in. end to end. The times to traverse the last 6 in. beyond the exit of the start box and each of the next seven 2-ft. sections were recorded by means of electronic switches, activated by contact with the floor grids, operating .01-sec. Standard Electric timers. The time spent in retracing was included in the time to traverse the last section reached before retracing began. Thus eight times were obtained, a starting time and the times taken to traverse the seven successive sections preceding the end box. The start-box door opened 3 sec. after the start-box lid closed. The end-box door came down when *S* touched the ninth grid, 6 in. beyond the entrance to the end box.

The magazine consisted of a 50-ml. burette to which a glass drinking tube bent at an angle of 120° had been annealed. The tip of the drinking tube was presented ½ in. behind a 1-in. diameter hole 2 in. above the floor in the rear wall of the end box. Two sizes of tips were used, 1 mm. and 4 mm. (internal diameter). Contact with the drinking tube activated an electronic switch which recorded the number of licks. The end box was illuminated by a 7-w. light located behind the hole for the tube.

The two concentrations (Percentage × Weight) of sucrose used were mixed from

[1] This investigation was supported in part by Research Grant M-3328 from the National Institutes of Health, Bethesda, Maryland. The computation in the analysis was performed by the University of Missouri IBM center and was supported by the Faculty Research Council. Experiment 1 is the first author's Master of Arts thesis. A brief report of Exp. 1 was given at the Midwestern Psychological Association meeting in St. Louis in 1960. Experiment 2 was performed by Earl Gates.

commercial sugar and distilled water 24 hr. before use.

Procedure.—Three volumes per reinforcement (.5, 1.0, and 1.5 ml.), two concentrations (8 and 32%), and two tube size openings (1 and 4 mm.), were combined factorially.

The *S*s first received 10 days of deprivation accommodation, handling, and table training. A 22-hr. food and water deprivation schedule was used. Each *S* was handled for 3 min. per day followed by 5 min. of table top exploration. On the last 3 of the 10 days, *S*s were placed for 10 min. in the runway with the doors up and no reinforcement present.

Following the table training and handling *S*s were shifted to a 22-hr. food deprivation and an ad lib. water schedule. Ten days of magazine training were given, consisting of 3 days of drinking the appropriate concentration from the approprite burette in the home cage and then 7 days of drinking one trial per day in the end box.

Finally, each *S* was given 20 runway trials, 1 per day, on 20 consecutive days. The rat was removed from the end box immediately following his consumption of the appropriate amount of solution. Any *S* who failed to traverse the runway within 10 min. on 2 consecutive trials was discarded.

TABLE 1

ANALYSIS OF VARIANCE OF STARTING SPEEDS, EXP. 1

Source	*df*	*MS*	*F*
Ss_b	83	.175	
Concentration (C)	1	6.72	73.16**
Tube size (T)	1	0	
Volume (V)	2	.032	.35
C × T	1	.011	.12
C × V	2	.220	2.39
T × V	2	.135	1.47
C×T×V	2	.189	2.06
$Error_b$	72	.092	
Ss_w	756	.021	
Blocks (B)	1	1.969	37.11**
Trials (R)	4	.102	8.81**
B×R	4	.052	5.45**
C×B	1	1.765	33.26**
C×B×R	4	.03850	4.06**
C×T×B×R	4	.035	3.69**
C×T×V×B×R	8	.007	.71
$Error_w$	648	.015	
E_1	72	.053	
E_2	288	.012	
E_3	288	.009	

** $P \leq .01$.

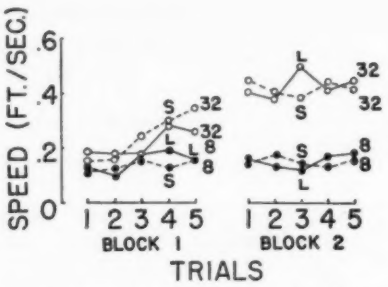

Fig. 1. Starting speed as a function of concentration, tube size, trial block, and trial.

Experiment 2

Subjects.—The *S*s were 15 female albino rats from the University of Missouri colony, 90–115 days old.

Apparatus.—An earlier version of the apparatus described above was used. The major difference was that the doors were solenoid operated. Noyes 45-mg. pellets were given as reinforcement in a glass coaster placed on the floor of the end box.

Procedure.—All *S*s received 7 days of deprivation accommodation to a 23-hr. schedule and 3 days of runway accommodation. The latter consisted of being placed 2 at a time for 6 min. in the runway in which Noyes pellets had been scattered. Following accommodation each *S* was given 35 trials, 1 trial per day.

Three magnitudes of reinforcement were used. The first group received 1 pellet and remained in the end box for approximately 2 sec. The second group received 8 pellets and remained in the end box for approximately 15 sec. The third group received 15 pellets and remained in the end box for approximately 30 sec. The end-box time represents the approximate time it took for *S* to consume the number of pellets offered as determined in preliminary trials on other rats. All *S*s were removed as soon as possible following consumption of their allotments. Fifteen pellets were used as the largest value since the preliminary experience showed that this was the maximum number of pellets that *S*s would consume without interruption.

RESULTS

Experiment 1

For purposes of analysis the first five and the last five trials were con-

sidered. All times were converted to speeds.

Table 1 presents an analysis of the starting speeds for concentration (C), tube size (T), volume per reinforcement (V), first and last block of trials (B), and trials (R). Figure 1 presents the highest order significant interaction, C × T × B × R. It is clear from Fig. 1 and the analysis that the 32% *Ss* started faster; that they improved over trials and trial blocks while the 8% *Ss* showed little if any improvement in starting speed over

TABLE 2

ANALYSIS OF VARIANCE OF RUNNING SPEEDS, EXP. I

Source	*df*	*MS*	*F*
Ss_b	83	67.76	
Concentration (C)	1	2400.34	61.41**
Tube size (T)	1	35.55	.91
Volume (V)	2	12.45	.32
C × T	1	11.66	.30
C × V	2	47.96	1.23
T×V	2	71.94	1.84
C×T×V	2	48.56	1.24
$Error_b$	72	39.09	
Ss_w	5796	1.52	
Gradient (G)	6	60.59	61.20**
Blocks (B)	1	2854.55	444.63**
Trials (R)	4	106.36	82.45**
G×B	6	10.18	16.94**
G×R	24	2.65	5.76**
B×R	4	84.21	63.32**
C×G	6	11.21	11.32**
C×B	1	345.37	53.80**
C×R	4	13.77	10.67**
C×V×R	8	3.99	3.09**
C×G×B	6	6.47	10.77**
C×G×R	24	.71	1.53*
C×B×R	4	24.92	18.74**
G×B×R	24	2.23	1.64*
C×T×G×R	24	.72	1.55*
C×T×V×G ×B×R	48	.40	.29
$Error_w$	4968	.72	
E_1	432	.99	
E_2	72	6.42	
E_3	288	1.29	
E_4	432	.60	
E_5	1728	.46	
E_6	288	1.33	
E_7	1728	1.36	

* $P \leq .05$.
** $P \leq .01$.

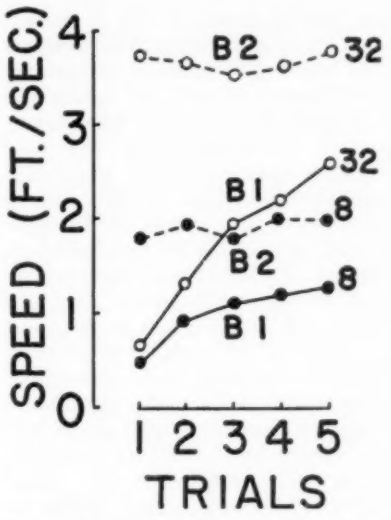

FIG. 2. Running speed as a function of concentration, trial block, and trial.

either trials or trial blocks; and that finally the large tube was superior for the 8% *Ss* and the small tube for the 32% *Ss* in the first block of trials and that there were no differences produced by tube size in the final block of trials. There was no evidence that volume per reinforcement had any significant effect on starting speed. Both the 8% and the 32% groups appeared to have reached a stable level of starting speed by 20 trials.

Table 2 presents an analysis of variance on running speed over C, T, V, B, R, and the sections of the runway, 2–8 (G). Figure 2 presents the C × B × R interaction. In the first block of trials the 32% group showed the most rapid learning, and in the second block of trials their performance was relatively stable while the 8% group was still improving. The final levels of performance of the 32% group were substantially greater than those of the 8% group. The first

and second panels of Fig. 3 present the C × B × G interaction. It shows that the initial gradient of running speed was a slightly increasing one and the final gradient was a decreasing one. The steepness of the final gradient was a function of concentration, the 32% group showed a sharper initial rise and a steeper decline than the 8% group. The first-trial gradient was flat. To show the evolution of the gradient beyond 20 trials the third and fourth panels depict the gradient of running from another experiment (Collier et al., 1961) in the same apparatus on Trials 42 and 64, respectively. The initial increasing gradient in running speed over the first portion of the runway was replaced by an initial burst with a continuous decline in speed over the remainder of the runway. The remaining two interactions of interest were the C × V × R and C × T × G × R. The first resulted from the fact that for the 32% group, running speed ranked high to low for the 1.0, .5, to 1.5 ml. volumes per reinforcement, while for the 8% group the high to low running speed ranking was .5, 1.5, to 1.0. These differences increased slightly over trials. The differences in running speed resulting from the volume differences were relatively small. The second interaction, involving tube size, was the result of a small but persistent tendency on the early trials and the first portion of the gradient for the *small* tube size opening to produce faster running in the 32% group, in contrast to a larger, consistent tendency on the first portion of the gradient for the *large* tube size opening to produce faster running in the 8% group. These results are similar to those on starting times. Again, the differences produced by tube size were small in comparison with those produced by concentration. Tube size had its greatest effect at the low concentration.

Table 3 presents the average number of licks, the volume per lick, and the running and starting speed in each

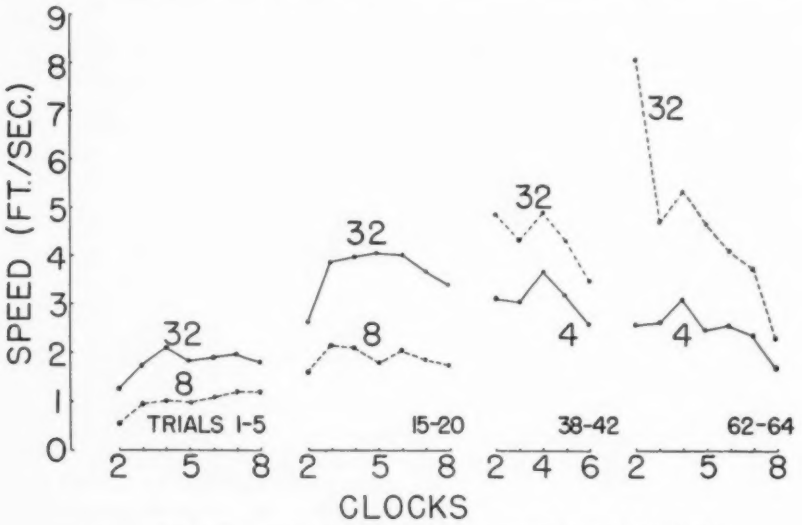

FIG. 3. Gradient of running as a function of concentration and trial block.

TABLE 3

LICKS, VOLUME PER LICK, STARTING, AND RUNNING SPEEDS AS A FUNCTION OF VOLUME PER REINFORCEMENT, CONCENTRATION, AND TUBE SIZE

Concentration %	Tube Size (Mm.)	Volume per Reinforcement (Ml.)	Licks	Volume per Lick (Ml.)	Starting Speed (Sec.)	Running Speed (Sec.)
8	1	.5	113.8	.0044	.13	1.58
		1.0	189.6	.0053	.14	1.85
		1.5	315.3	.0048	.18	1.81
8	4	.5	55.4	.0090	.11	1.96
		1.0	120.7	.0083	.19	2.22
		1.5	172.0	.0087	.14	2.03
32	1	.5	125.6	.0040	.39	3.39
		1.0	241.3	.0041	.43	3.60
		1.5	349.8	.0043	.43	3.90
32	4	.5	58.3	.0086	.48	4.21
		1.0	139.4	.0072	.29	3.08
		1.5	177.7	.0084	.50	3.85

group. An analysis of variance of the lick data for C, T, and V showed C, $P \leq .01$ ($F = 17.80$; $df = 1/72$); T, $P \leq .01$ ($F = 424.22$; $df = 2/72$); V, $P \leq .01$ ($F = 372.15$; $df = 2/72$); C×T, $P \leq .05$ ($F = 5.64$; $df = 1/72$); and T × V, $P \leq .01$ ($F = 33.40$; $df = 2/72$) reaching significance. The 32% solutions required a significantly greater number of licks to get the offered volume which is reflected in the smaller volumes per lick at the higher concentration (Table 3). As would be expected, the small tube required a larger number of licks to obtain the required amount of solution. Tube size interacted with concentration; the difference between tube sizes was greatest at 32%. Volume and tube size also interacted, the largest volume per lick occurred at the large volume for the small tube size while the largest volume per lick occurred at the small volume for the large tube size.

The rank correlation between number of licks and starting speed was .35 ($P > .05$) for the 12 pairs of comparisons. The correlation between running speed and licks was .03 ($P > .05$). The correlation between starting and running speed was .72 ($P \leq .01$).

In summary, the rate of acquisition, the final level of performance, and the gradient of running were all functions of the concentration of reinforcement. By cross comparisons between the various volumes per reinforcement and tube size conditions it can be seen that these differences can not be attributed either to the number of licks, the volume consumed, or the duration of activity; i.e., the largest number of licks, the largest volume per reinforcement, or the longest time were not consistently associated with the fastest running. In fact, the fastest running time and the second fastest starting time are associated with next to the smallest number of licks.

The initial gradient of running for premagazine trained *Ss* showed an initial rise and was relatively flat throughout the remaining distance. The final gradient showed an initial burst of speed with a progressive decline over the length of the runway.

TABLE 4

STARTING SPEED (FT/SEC) ON THE FIRST AND LAST BLOCK OF TRIALS

Number of Pellets	Blocks	
	First	Last
1	.166	.391
8	.214	.846
15	.230	.830

Experiment 2

Table 4 presents the average starting speeds for the three groups on the first and last block of the seven blocks of trials. An analysis of variance, Table 5, shows that both number of pellets and trial blocks were significant. The 15-pellet *S*s ran faster than the 8 and the 8 faster than the 1. All groups improved in starting speed over trials. There was no significant interaction.

Figure 4 shows the running speed over the various segments of the runway for the average of the first five trials and the average of the last five trials. An analysis of the data is presented in Table 6. The initial gradient is flat. The subsequent evolution is similar to that reported in Exp. 1, showing in its terminal phase, the fastest running in the first segment with a gradual decline

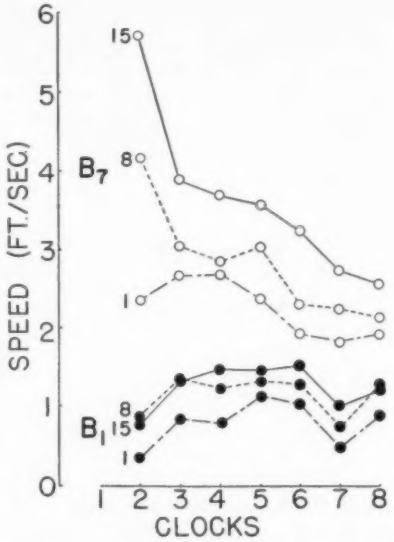

FIG. 4. Gradient of running as a function of number of pellets and trial block.

TABLE 5

ANALYSIS OF VARIANCE OF STARTING SPEEDS IN EXP. 2

Source	*df*	*MS*	*F*
Ss_b	14		
Number (N)	2	1.5069	3.88*
$Error_b$	12	.3880	
Ss_w	90		
Blocks (B)	6	.4629	12.25**
N × B	12	.0653	1.73
$Error_w$	72	.0378	

* $P \leqq .05$.
** $P \leqq .01$.

in running speed in the direction of the end box. On the intermediate trials the fastest running occurred in the middle portion of the runway. The failure of N or any of its interactions to be significant appears to be the result of a single deviant *S* in the 1-pellet group who showed the fastest running and the steepest gradient of any *S*. When this *S* is dropped from the analysis and the group average in each block of trials at each segment is substituted, all of the interactions containing N prove significant at at least the .05 level. An examination of the individual curves for the 15-pellet group revealed that every *S* showed the gradient while only the deviant *S* in the 1-pellet group showed a consistent trend.

Figure 5 presents the speed of running for starting speed, and for the first and last segment of the runway plotted as a function of the number

of pellets offered. The first trial curve is flat. In summary, the asymptotic gradient of running is negative and its steepness is a function of the magnitude of reinforcement. The speed of starting and running increases linearly with the number of pellets offered.

Discussion

Amount of reinforcement.—It is clear from this and preceding studies (e.g., Collier et al., 1961; Kling, 1956; Stein, 1957) that variations in total number, average rate, or duration of licking, and volume or average volume consumed per second, are neither necessary nor even sufficient conditions for variation in acquisition or performance.

The studies of Stellar and Hill (1952), and Davis and Keehn (1959) show further that the momentary rates of licking or consumption are constant under a wide variety of conditions and, therefore, can not be considered determinants of the rate of acquisition or the level of performance.

An alternative account of the often found association between measures of consummatory responding and instrumental performance is that they are independent functions of some of the same variables. If we assume that the consummatory responses are respondents, whose strengths are functions of the intensive properties of the eliciting stimuli, and that one class of determinants of the amount of reinforcement function is the sensory consequences of the reinforcing substance (e.g., Collier & Myers, 1961), then such covariation would be expected. However, when one of these functions, for example amount of reinforcement as a function of volume, approaches an asymptote and the other does not, or when a given variable, for example tube size, affects one of these measures and not the other, then such covariation would not be expected. In those cases (e.g., Goodrich, 1960; Hellyer, 1953) where manipulation of the amount of consummatory activity appears to be a determinant of magnitude of reinforcement, it is possible that it gains this effect via the contribution of the consummatory responses to the sensory pool as well as the volume consumed per reinforcement.

TABLE 6

Analysis of Variance of Running Speeds in Exp. 2

Source	*df*	*MS*	*F*
Ss_b	14		
Number (N)	2	55.1173	1.59
$Error_b$	12	34.7511	
Ss_w	720		
Clocks (C)	6	9.7751	14.49**
Blocks (B)	6	45.3567	52.92**
B × C	36	1.3413	5.66**
N × B	12	1.3852	1.62
N × C	12	.8125	1.20
N × B × C	72	.2462	1.04
$Error_w$	576		
E_1	72	.6747	
E_2	72	.8571	
E_3	432	.2373	

** $P \leq .01$.

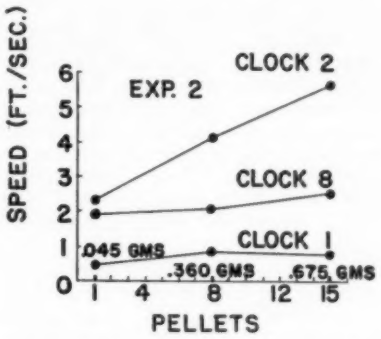

Fig. 5. Runway performance as a function of number of pellets.

Gradient.—The model of runway behavior that views it as a sequence of discrete responses pinned together by stimuli arising from the runway, from the deprivation operation, and from fragments of the consummatory response (e.g., Hull, 1932) implies a positive gradient of running, at least in the early acquisition trials. In spite of the textbook certainty and its appealing

elegance, few examples of a "goal gradient" in the linear runway have been reported. These examples of positive gradients seem to occur under three restricted conditions: (*a*) Animals which have not been adapted to the runway typically show a positive gradient on early trials (e.g., Bruce, 1937; Drew, 1939; Epstein & Morgan, 1943; Hull, 1934). Pereboom (1958) has shown a positive gradient can exist on the first trial. (*b*) Runways with obstacles, such as the doors used by Hull (1934) and Drew (1939) for timing, choice points, running wheels (e.g., Meade & Newman, 1960), and delay doors (e.g., Wist, 1962) produce a positive gradient, at least in early trials. (*c*) Low deprivation (Bruce, 1937; Morgan & Fields, 1938) and small magnitudes of reinforcement (Crespi, 1942; and present study) produce a positive gradient. A further condition, which has not been systematically examined, producing a tendency toward a positive gradient is the use of open runways where there is a large amount of extraneous stimulation.

However, after extended training in a linear unobstructed runway, with a modicum of deprivation and reinforcement, the most characteristic gradient of speed of running is negative (e.g., Bruce, 1937; Collier et al., 1961; Crespi, 1942; Morgan & Fields, 1938; Weiss, 1960; and present study). The final gradient is a monotonically decreasing one with the fastest running occurring in the initial portion of the runway. The steepness of this final gradient is subject to such variables as amount of reinforcement (Crespi, 1942; Weiss, 1960; and present study), deprivation (Weiss, 1960), length of runway (Besch & Reynolds, 1958), and delay. The gradient, once obtained, shifts relatively rapidly with shifts in amount of reinforcement (Collier et al., 1961) or delay or position of delay (Wist, 1962). Thus it would seem that in its present form, at least, the goal gradient hypothesis can not account for the distribution of speeds of running in the runway.

An alternative account can be derived from the assumption that two classes of chaining of responses exist. The first class contains repetitive serial reflexes of a fixed, genetically determined topography such as licking, chewing, running, etc., in which the chaining stimuli are those produced by the preceding response. These sequences are not chained by external stimuli. They can be brought under the control of specific discriminative stimuli such as presentation of food, opening of a start-box door, etc. The second class contains responses of novel flexible evolutionary topography, such as right and left turns in a maze, sequential discrimination behavior, etc., which are chained by external stimuli. There exists also an anomolous middle class which has as members responses of the second class which are repetitive such as bar pressing and for which the chain shifts from external to internal stimulus control as learning proceeds. These two classes of chains have quite different properties, among which are the fact that the first acts as a unit, i.e., occurs in bursts of constant rate (e.g., Premack & Schaeffer, 1962; Stellar & Hill, 1952), the appropriate parameters being burst rate, burst duration, and burst latency; while for the second the individual response acts as a unit, the appropriate parameters being the probability and/or latency of occurrence. It is only for this latter class of chaining that the classical gradient of reinforcement would be expected.

If now runway behavior is considered as consisting of three independent classes of behavior, starting, running, and stopping, an account of the relations obtained can be given. Starting is a response elicited by an external stimulus and is highly sensitive to consequences, reinforcement, delay, etc., of the trip down the runway. Running is an internally chained response which occurs at a constant rate in bursts of varying duration and intervals between bursts. Differences in average running speed are the result of variations in these parameters of running. Premack and Schaeffer (1962) have further suggested that discrete burst rates exist

as a function of the rat's gait. Thus most of the running in a linear runway is not under the control of an external sequence of stimuli but rather it consists of an internally chained sequence of responses of which only the initial member may be sensitive to the conditions of reinforcement, etc. The degree to which responses elicited by the nondiscriminative stimuli in the runway compete with starting and running appear to be a function of intra- and intertrial adaptation (e.g., Hull, 1934), amount of reinforcement (e.g., Armus, 1959; Pereboom & Crawford, 1958; and present study), and deprivation (e.g., Cotton, 1953). Burst duration and interval between bursts are functions of the parameters of reinforcement, deprivation, etc. The occurrence of a gradient of running may be attributed to the distribution of burst rates and burst durations. The probability of a fast burst is greatest in the initial portion of the runway since its initial member is starting, which is most sensitive to the conditions of reinforcement. A positive gradient of running is associated with a slow start however produced and a negative gradient with a fast start however produced. Such a relation can be observed in the individual trials where 72% of the block of last five trials for the 5 *Ss* in the 15-pellet group of Exp. 2 were monotonic, that is had their fastest run in the first segment. Ninety-two percent had their fastest run in either the first or second segment. It seems likely that the shape of the gradient is the result of phenomena occurring in the start box rather than in the end box or in the rat's mouth.

Summary

In two runway experiments the effects of jointly manipulating concentration of sucrose, volume per reinforcement, and amount of consummatory activity on starting and running speed were examined. Rate of acquisition and the terminal level of responding proved to be functions mainly of concentration and of number of pellets offered. The relatively minor effects of tube size occurred in the early training trials and those of volume per reinforcement in the terminal trials.

It was concluded that variations in consummatory activity are neither necessary nor even sufficient conditions for producing variations in acquisition and performance.

The implications of the fact that the terminal gradient of running was a negative one whose steepness was a function of concentration and amount were explored.

References

Armus, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58,** 61–63.

Besch, N. F., & Reynolds, W. F. Alley length and time of food deprivation in instrumental reward learning. *J. exp. Psychol.*, 1958, **56,** 448–452.

Bruce, R. H. An experimental investigation of the thirst drive in rats with especial reference to the goal gradient hypothesis. *J. gen. Psychol.*, 1937, **17,** 49–62.

Collier, G., Knarr, F., & Marx, M. Some relations between the intensive properties of the consummatory response and reinforcement. *J. exp. Psychol.*, 1961, **62,** 484–495.

Collier, G., & Myers, L. The loci of reinforcement. *J. exp. Psychol.*, 1961, **61,** 57–66.

Cotton, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46,** 188–198.

Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55,** 467–517.

Davis, J. D., & Keehn, J. D. Magnitude of reinforcement and consummatory behavior. *Science*, 1959, **130,** 269–271.

Drew, G. C. The speed of locomotion gradient and its relation to the goal gradient. *J. comp. Psychol.*, 1939, **27,** 333–372.

Epstein, M. A., & Morgan, C. T. Cortical localization of symbolic processes in the rat: III. Impairment of anticipatory functions in prefrontal lobectomy in rats. *J. exp. Psychol.*, 1943, **32,** 453–463.

Goodrich, K. P. Running speed and drinking rate as functions of sucrose concentration and amount of consummatory activity. *J. comp. physiol. Psychol.*, 1960, **53,** 245–250.

Hellyer, S. Duration of the consummatory response as a variable in amount of reinforcement studies. Unpublished doctoral dissertation, Indiana University, 1953.

HULL, C. L. The goal gradient hypothesis and maze learning. *Psychol. Rev.*, 1932, **31,** 25–43.

HULL, C. L. The rat's speed of locomotion gradient in the approach to food. *J. comp. Psychol.*, 1934, **17,** 393–422.

KLING, J. W. Speed of running as a function of goal-box behavior. *J. comp. physiol. Psychol.*, 1956, **49,** 474–476.

MEADE, R. D., & NEWMAN, P. J. Locomotion gradients in the activity wheel. *J. comp. physiol. Psychol.*, 1960, **53,** 381–384.

MORGAN, C. T., & FIELDS, P. E. The effect of variable preliminary feeding upon the rat's speed-of-locomotion. *J. comp. Psychol.*, 1938, **26,** 331–348.

PEREBOOM, A. C. A goalless gradient. *J. exp. Psychol.*, 1958, **55,** 31–33.

PEREBOOM, A. C., & CRAWFORD, B. M. Instrumental and competing behavior as a function of trials and reward magnitude. *J. exp. Psychol.*, 1958, **56,** 82–85.

PREMACK, D., & SCHAEFFER, R. W. Distributional properties of operant-level locomotion in the rat. *J. exp. Anal. Behav.*, 1962, **5,** 89–95.

STEIN, L. The classical conditioning of the consummatory response as a determinant of instrumental performance. *J. comp. physiol. Psychol.*, 1957, **50,** 269–278.

STELLAR, E., & HILL, J. H. The rat's rate of drinking as a function of water deprivation. *J. comp. physiol. Psychol.*, 1952, **45,** 96–102.

WEISS, R. F. Deprivation and reward magnitude effects on speed throughout the goal gradient. *J. exp. Psychol.*, 1960, **60,** 384–390.

WIST, E. R. Amount, delay, and position of delay of reinforcement as parameters of runway performance. *J. exp. Psychol.*, 1962, **63,** 160–166.

(Received April 28, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 589–592

ACTIVITY DURING DELAY OF REINFORCEMENT IN HUMAN LEARNING

R. A. CHAMPION AND D. A. McBRIDE

University of Sydney

Compared with immediate reward, delayed reward leads to inferior performance on the part of subhuman animals in learning situations. A number of attempts have been made to explain this phenomenon, the most recent being that of Spence (1956), who has supposed that a delay period allows the stimulus components to become attached to a variety of responses which compete with the response being learned, so that the longer the delay the greater the number of competing responses and the poorer the performance. This hypothesis was supported in a study by Carlton (1954) in which the effect of delay of food reward was significantly reduced in rats when competing responses were kept to a minimum by confinement during the delay period. On the other hand, a series of experiments on the effect of delay of knowledge of results in human learning (e.g., Bilodeau & Ryan, 1960) has revealed no comparable interference with performance, and some resolution of this apparent conflict in results is required. The demand which led to the following study has been explicitly stated by Noble and Alcock (1958) as an examination of human learning over a range of delay intervals both with and without disrupting interpolated activity. The present experiment was accordingly designed to serve two purposes: (*a*) a further test of Spence's hypothesis, and (*b*) an exploration of the possible application of the hypothesis to human performance.

With regard to the first of these purposes, Carlton had to be content with the assumption that the unconfined rats would be active during delay and in fact observed that some such animals were not active. In order to achieve a higher degree of control over activity during the delay interval, rats could be presented with a stimulus evoking a previously learned response or human *S*s could be made active under instruction; the latter alternative was chosen in the present case. This choice having been made, care was taken that Spence's hypothesis might be meaningfully applied to human *S*s, in keeping with the second purpose of the experiment. The effects of delay of reward in lower animals have mostly been obtained with latency or speed as a measure, applied to an act already high in the response hierarchy (such as lever pressing). The human *S*s were therefore simply required to give the response items for very easy paired associates and no measure of performance was taken unless the response was correct. A verbal response was preferred to a motor act because it allowed the presentation of the correct response item to be used as reinforcement. It should be noted that the term "reinforcement" is being used here in a purely empirical sense (Champion, 1960), for even performance with easy paired associates presumably changes little unless the response is followed by the correct item.

METHOD

Subjects.—There were 80 *S*s, recruited mainly from undergraduate classes at the

University of Sydney. The only requirements for selection were that *S* be free from speech defect and naive as to the purpose of the experiment.

Apparatus.—The material to be learned consisted of five paired associates as follows: NEEDLE-THREAD, SWEET-SOUR, TABLE-CHAIR, SHEEP-ANIMAL, STEM-FLOWER. These were made up on the basis of the Kent-Rosanoff study of free association (Rosanoff, 1938) such that the response in each case was the most frequent association and no stimulus had been given as a response to any of the other stimuli. The items were presented in a Ranschburg apparatus and constant exposure periods were provided by electronic timers. The latency of *S*'s response to each stimulus word was measured with an electronic chronoscope which was automatically started with the presentation of the stimulus and stopped by *E* as soon as *S* responded.

Two groups of *S*s were required to read words aloud during the period between their response and the appearance of the response word, and the interpolated words were also common associations with the stimulus word in the Kent-Rosanoff study. For example, the words to be read out after the response to the stimulus NEEDLE were: SHARP, PIN, SEW, STEEL, EYE, INSTRUMENT, POINT, THIMBLE, SEWING. These items were typewritten in groups according to the stimulus word on stiff 5 × 2½ in. cards hinged in an aluminum case, with tabs to facilitate the handling of the cards, and were placed at *S*'s right hand; the top two cards were blank. The paired associates and the corresponding sets of words to be read aloud were presented in three different orders, and the sequence of the latter words for any one pair was varied in each order.

Procedure.—Two periods of delay of reinforcement were used, 2 sec. and 5 sec.; half the *S*s on each delay period were required to read words aloud during delay whereas the remaining *S*s were instructed to fixate the slot in the exposure apparatus. This factorial design gave four groups designated henceforth as follows: (*a*) Group 2N: 2-sec. delay of reinforcement with no activity during delay, (*b*) Group 2A: 2-sec. delay with activity, (*c*) Group 5N: 5-sec. delay with no activity, and (*d*) Group 5A: 5-sec. delay with activity. The *S*s were allocated to these groups in successive cycles according to the order of their appearance at the laboratory, giving 20 *S*s in each group.

Each presentation began with *E* giving a verbal ready signal, followed about 1 sec. later by the activation of the exposure apparatus to present the stimulus word and to start the chronoscope. Immediately as *S* responded *E* simultaneously stopped the chronoscope, reactivated the exposure apparatus to present a blank to *S*, and started the timer controlling the delay-of-reinforcement interval. After 2 sec. or 5 sec. the exposure apparatus was automatically activated once more so as to present the response word for 1 sec., followed automatically by another blank. One run through the list of five items was regarded as a trial, and training was continued for 15 trials (75 presentations). The interval between presentations averaged approximately 20 sec.

Instructions.—The *S*s in Groups 2A and 5A were instructed as follows:

> This experiment is concerned with the learning of certain words as responses to other words. When a word appears in the slot in front of you, say the first word you think of. A few seconds after you give your reply a second word will appear and this is the response you are to learn. At first you will simply be guessing but the task is easy and you will learn it quite quickly. After you give your response, but before the correct response appears, you are to flip down the blank card thus (at this point the first blank card was pulled down as a demonstration) and read aloud the words which are typewritten on the card behind it. You are to pull down a card and start reading immediately after you make your response each time. Read these words aloud at the rate of about one per second. Stop reading as soon as the correct response word clicks into the slot. If you should finish reading the words on a card before the correct word has appeared, start reading at the beginning of the same card again. Remember that although the correct responses are easy to learn it is the speed with which you give your response each time that counts. I will say "now" about 1 sec. before each word to which you are to respond appears in the slot. Have you any questions?

After 15 responses had been given, so that all the word lists had been read from the cards, the *S*s in Groups 2A and 5A were instructed to push all the cards except the first blank back into their original position and to continue to do this each time all the cards had been read. The *S*s in Groups 2N and 5N were given the same general instructions as the other *S*s except that all references to reading words aloud were eliminated and

they were instead asked to watch the exposure slot carefully throughout each trial, especially when waiting for the correct response word to appear.

Results

Each raw latency score in seconds was converted to a speed measure by a reciprocal transformation and the reciprocals were averaged over successive trials (sets of five responses). The time measures for wrong responses were discarded, but there were very few of these after the first trial. Performance curves drawn from these data are shown in Fig. 1 and the results of an analysis of variance are set out in Table 1. The three main effects of practice, activity, and delay were significant, along with the double interactions of practice with activity and delay. The mean score for the first response in each group was included in Fig. 1 as a check on the comparability of the groups at the outset of training.

Discussion

Under the highly circumscribed conditions of this experiment with human *Ss* results have been obtained which tally with those of Carlton's (1954) study using rats; since activity during delay of reinforcement impaired performance then Spence's hypothesis may be said to have been confirmed again. It might be argued that a strict application of the competing-response hypothesis should imply no impairment of performance with no activity during delay of reinforcement, so that Groups 2N and 5N should not have differed. The obtained superiority of Group 2N, however, may be ascribed to the technical impossibility of rendering *S* completely inactive, even for brief intervals (Bilodeau & Ryan, 1960), so that the designations A and N in this study refer to high and low degrees of activity rather than to its presence and absence, respectively. A further point of possible criticism concerns the interval between presentations. No attempt was made to hold this interval constant and it probably averaged 3 sec. longer for the 5-sec. delay groups. Bilodeau and Bilodeau (1958) have found some evidence for superior performance with shorter intertrial intervals as compared with a 7-day interval, but they found no significant differences with intervals of 20 sec., 2 min., and 1 hr. Denny and associates (Denny, Allard, Hall, & Rokeach, 1960) have reported significant differences with intervals of 10 sec. vs. 30 sec. and 20 sec. vs. 30 sec., but these *Es* applied no over-

TABLE 1

Analysis of Variance of Mean Response Speeds

Source	*df*	*MS*	*F*
Between *Ss*	79		
Degree of activity (B)	1	9409.00	21.92**
Delay interval (C)	1	3306.25	7.70*
B × C	1	104.04	
Error	76	429.15	
Within *Ss*	320		
Trials (A)	4	7735.61	369.77**
A × B	4	148.05	7.08**
A × C	4	108.77	5.20**
A × B × C	4	8.71	
Error	304	20.92	
Total	399		

* $P < .01$.
** $P < .001$.

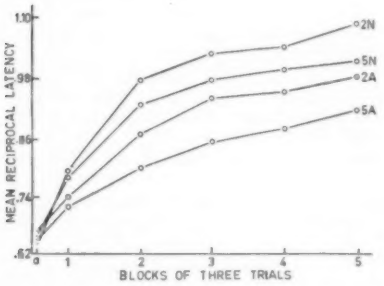

Fig. 1. Performance curves for groups learning with 2 and 5 sec. delay of reinforcement, and with activity (A) or inactivity (N) during the delay period. (The scores for Point a on the abscissa were taken from the the first item presented in Trial 1.)

all test of significance to their data before proceeding to compare groups in pairs. It thus seems reasonable to conclude that the possible difference between intervals in this study, of the order of 19 sec. vs. 22 sec., could have had no appreciable effect on performance.

In view of the positive results of the present study, despite the use of human *Ss*, the conflict with the null results of delayed knowledge of results remains a problem, but the apparent discrepancy may depend on a false analogy between the two situations. In the experiment under consideration and in previous similar work with lower animals, the reinforcing event cannot be regarded as providing knowledge of results, for there has been no question of the correctness of the response, only its strength; a methodological concomitant of this fact has been the study of simple responses and the use of reaction latency or speed as a sensitive measure of response strength. In the other experiments with human *Ss*, however, responses requiring some skill in their execution have been studied, the learner has been given some indication of the correctness of the response, and relatively crude or complex measures, such as frequency of correct response, have been used. The present *Es* sought to maintain the conditions of Spence's hypothesis with humans; if the process were to be reversed, so that rats served as *Ss* in the usual type of study manipulating knowledge of results, then rather different conditions should obtain. For example, the rat as *S* might be required to learn to exert a certain degree of pressure on the lever, amount of reward could be scaled to the proximity of the obtained degree of pressure to that required, and frequency of correct pressure or "error" of pressure would then presumably be used as measures of performance. Under this latter set of conditions it could well prove that delay of reward has no measurable effect with lower animals, as has been the case with humans.

Summary

With speed of response as a measure it was found that both delay of reinforcement and activity during delay had detrimental effects in a very simple learning situation with human *Ss*. This result confirmed data from a similar study using rats and was interpreted as further supporting Spence's (1956) hypothesis about delay effects in terms of competing responses. Experiments on delay of knowledge of results in human motor learning, yielding no such effects, were seen to differ in certain basic respects from the present type of study.

References

Bilodeau, E. A., & Bilodeau, I. McD. Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

Bilodeau, E. A., & Ryan, F. J. A test for interaction of delay of knowledge of results and two types of interpolated activity. *J. exp. Psychol.*, 1960, **59**, 414–419.

Carlton, P. L. Response strength as a function of delay of reward and physical confinement. Unpublished master's thesis, State University of Iowa, 1954.

Champion, R. A. Reinforcement and learning theory. *Aust. J. Psychol.*, 1960, **12**, 10–20.

Denny, M. R., Allard, M., Hall, E., & Rokeach, M. Supplementary report: Delay of knowledge of results, knowledge of task, and intertrial interval. *J. exp. Psychol.*, 1960, **60**, 327.

Noble, C. E., & Alcock, W. T. Human delayed-reward learning with different lengths of task. *J. exp. Psychol.*, 1958, **56**, 407–412.

Rosanoff, A. J. *Manual of psychiatry.* N. Y.: Wiley, 1938.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received May 1, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 593–600

THE DISTRIBUTION OF RECALLED ITEMS IN SIMULTANEOUS INTENTIONAL AND INCIDENTAL LEARNING[1]

ARNOLD MECHANIC[2]

University of California, Berkeley

Two types of incidental learning situations may be distinguished: Type I in which *S* is not instructed to learn any of the materials but is later tested; Type II in which *S* is given a circumscribed learning task and is later tested for materials to which he was exposed during learning but which are not covered by the learning instructions. In both situations, *S* is introduced to the incidental stimulus materials by means of an orienting task. This is a task which *S* is required to perform in order to insure that he is exposed to the incidental stimuli. The activities required by differing orienting tasks may vary widely as long as they insure *S*'s exposure to the incidental materials.

If instructions to learn generalize to materials not included in the instructions, amount of incidental learning should be greater in the Type II than in the Type I situation. However, such effects of generalization were not found in two previous studies using nonsense syllables. On the contrary, these studies, which will be reported in a forthcoming publication, indicated that the Type I situation produces significantly greater incidental learning. This was found to be true regardless of whether the intentional and incidental items were presented simultaneously or in temporal succession. If there were positive effects of generalization of instructions in these experiments, it would seem that they were completely countered by the negative effects of "task competition" in the Type II situation. By "task competition" we refer to the fact that the two kinds of material —the prescribed and the incidental— compete in terms of how much of the total exposure time is spent in responding to each kind of item. When *S*s are instructed to learn specific items, more of the available time will be utilized in responding to these items during exposure and in possible rehearsal thereafter. Consequently less time (or possibly no time) will be available for responding to the incidental items. The experiment to be reported in this paper studied task competition by manipulating the difficulty (meaningfulness) of both the intentional and incidental materials in a Type II situation. In addition, the effects of number of presentations were studied.

A review of the literature indicates no verbal learning data on Type II incidental learning as a function of intentional task variables. However, studies with nonverbal

[1] Based upon a thesis submitted to the Graduate School of the University of California, Berkeley, in partial fulfillment of the requirements for the PhD degree. The author is greatly indebted to his major professor, L. J. Postman, under whose direction this research was carried out and to B. J. Underwood for his suggestions in the preparation of this manuscript. This research was conducted during the author's tenure on a research fellowship (MF-9934) from the National Institute of Mental Health of the United States Public Health Service.

The manuscript was prepared during the author's tenure as a National Science Foundation Postdoctoral Fellow at Northwestern University.

[2] Now at Alameda State College, Hayward, California.

materials indicate that Type II learning varies with manipulations of the intentional task. Bahrick (1954) and Bahrick, Fitts, and Rankin (1952) found that the amount of learning irrelevant to a set is inversely related to the strength of the incentive determining that set. It might, therefore, be expected that varying the difficulty of the intentional items should also make for variations in the competition between the intentional and incidental tasks. Difficult intentional items could be expected to take up different proportions of *S*'s total learning time than would easy intentional items.

Both meaningfulness and degree of practice have been studied in the Type I situation. The studies varying meaningfulness indicate that the superiority of intentional over incidental learning decreases as the meaningfulness of the items increases (Postman, Adams, & Phillips, 1955; Postman & Phillips, 1961). These findings were explained in terms of the differential responses which *S*s make to the stimulus items. More specifically, differential responses may be regarded as habits attached to a stimulus item either through previous experience with that item or through stimulus generalization. It is assumed that learning—both intentional and incidental—depends on the performance of these differential responses to the stimuli during exposure. In this view, the major effect of learning instructions is to increase the frequency and intensity of such differential responses. It may also be assumed that the high-meaningful items will evoke strong differential responses at the outset, while the low-meaningful items will not readily evoke such responses.

As a result of the strong pre-experimental habits attached to them, many of the high-M items will be learned in the absence of learning instructions. In other words, these items should be learned quite easily by both incidental and intentional learners. The low-M items, on the other hand, could be expected to evoke only weak responses in the absence of learning instructions. Instructions to learn would be required in order to strengthen these responses to the degree that they are capable of mediating learning. Consequently, the lower the meaningfulness of the materials, the greater should be the effects of instructions to learn.

In the Type II situation, each *S* is exposed to both intentional and incidental lists. If the differential response hypothesis is correct, the difference between intentional and incidental scores for each *S* should decrease as the meaningfulness of the items increases.

The Type I data on number of presentations indicate that this variable has only marginal effects on the number of incidental items retained (Postman & Adams, 1958). Intentional learners showed a significant upward trend as a function of presentations whereas incidental learners did not. The present experiment will provide additional data on incidental learning as a function of presentations.

TABLE 1

STIMULUS LISTS

List A—Low	List B—Low	List C—High	List D—High
DUR	OAC	WHI	ANC
LAK	TOS	SHO	DRE
DOP	BAK	ATI	LAN
HOK	LDE	TES	GRA
FUM	DEL	OUN	ITY
HEO	LOA	HAN	ILD
ERF	KDE	RON	BLE
RHA	SYC	WER	UST
COF	WIC	UND	OVE
WYE	KAB	NES	TER
SIL	EJO	OME	OUS
ARP	MPL	EST	RES

METHOD

Materials.—Meaningfulness was defined in terms of language frequency of trigrams (three-letter sequences not forming words). The learning materials consisted of four 12-unit trigram lists. Two of these lists were of very high-frequency items while the other two were of very low-frequency items. Two of the lists, one high-frequency and one low-frequency, were taken from Underwood and Postman (1960). The other two lists were constructed so as to approximate the characteristics of the Underwood and Postman lists. Items were selected on the basis of the Underwood "total" frequency count. The manner in which these frequencies were derived is described by Underwood and Schulz (1960). The four lists are shown in Table 1.

There is very little overlap in frequency

between the low and high lists. On the basis of frequency per trigram in a sample of 1,035,000 English words, the frequency ranges are as follows: List A, 2–188; List B, 1–158; List C, 121–799; and List D, 127–1094. The approximate median frequencies are 5, 3, 350, and 375, respectively. None of the 48 trigrams appears in the Thorndike-Lorge (1944) word count.

The construction of these lists was governed by two criteria in addition to frequency. The first was formal intralist similarity defined in terms of duplication of letters. These lists were fairly well equated on this variable with the number of different letters for Lists A, B, C, and D being 17, 16, 13, and 16, respectively. The lower total number of letters used in List C was due to greater repetition of letters between first and second and between second and third positions.

The second criterion governing selection of items was the formal interlist similarity for all possible pairs of lists. Equating the pairs of lists on this variable was difficult because pairs of like frequency lists (H-H or L-L) tended to have more letter duplications than did pairs of mixed frequency lists (H-L or L-H). This is in all likelihood a function of different distributions of individual letters in the two populations (high-frequency trigrams vs. low-frequency trigrams). Therefore, the unmixed list pairings would be restricted to a narrower range of letters than would be the mixed pairings. Nevertheless it was possible to equate interlist similarity within reasonable limits. Duplicated letters between lists ranged from 9 to 13. Duplications in Position 1 ranged from 4 to 6; in Position 2, from 2 to 6; in Position 3, from 3 to 6. The unmixed list pairings averaged 5.5, 5.5, and 5 duplications in the three positions, respectively. The corresponding averages for the mixed list pairings were 5, 3.5, and 3.25.

Eight random orders were prepared for each of the four lists. Two of the lists were presented concurrently to each *S* during training. Each exposure of the lists involved presenting *S* with 12 pairs of items. Each pair was arranged vertically with one member in the top position and the other member directly below it. This arrangement was used to prevent the left to right chaining which might result from the reading habits of *Ss*.

The top members of the pairs made up one of the stimulus lists while the bottom members made up the second list. Successive presentations of the paired lists involved different random orders for each list. As a result, the pairings of individual items from the two lists were randomly varied from trial to trial. Stimulus materials were prepared for each of the six possible pairings of the four lists. Two variations for each pairing were prepared. These reversed the top and bottom member positions but were otherwise identical.

Conditions of learning.—All *Ss* were instructed to learn one of the two lists presented and were discouraged by the instructions from attempting to learn the second list. Different groups were exposed to each pair of lists for two, four, or eight presentations. The item pairings were varied from trial to trial in order to prevent learning of the incidental items through the mediation of the intentional items. The experiment was introduced as a test of the notion that words meaning the same thing in different languages tend to sound more alike than words with different meanings. The *Ss* were asked to judge the phonetic similarity between the members of each pair of trigrams. These were said to be items from two different primitive languages. The *Ss* were required to pronounce the members of each pair to themselves and to rate their phonetic similarity on a five-point scale. They were instructed that the top members of the pairs were always from the same language while the bottom members were always from the other language. In order to get intentional learning of only one list, *Ss* were told that we also wanted to study the ease of learning of one primitive language under conditions of distraction from a second language. In this way, instructions to learn were given for either the top or bottom items exclusively.

The *Ss* were given a number of paper sheets, each covered with a strip of cardboard. A window was cut out of each cardboard strip in order to allow exposure of one pair of syllables at a time in accordance with *E*'s instructions. The pairs of syllables on each sheet were numbered 1 through 12. When *E* called out the number "1," *S* slid down the window to expose the first pair of syllables and entered his rating next to the syllables. The procedure was repeated for each successive pair until all 12 pairs on the page had been rated. The pairs were presented at a 10-sec. rate, with 20 sec. between presentations of the pages. A 5-min. test of free recall was given 1 min. after the end of practice. The *Ss* were requested to write down as many of the syllables as they could remember regardless of whether they appeared in the top or bottom positions of the pairs. They were told to write the syllables down in the order they were remembered, and that their scores

TABLE 2

AVERAGE NUMBER OF INCIDENTAL ITEMS RECALLED BY THE VARIOUS GROUPS

Presentations	Incidental Learning Items			
	High Meaningful		Low Meaningful	
	High Int.	Low Int.	High Int.	Low Int.
2	2.31	2.25	1.06	1.44
4	3.56	4.06	2.31	2.62
8	4.94	4.69	3.88	3.19

would depend only on the number of correct items that they wrote down. When the recall test was completed, a postexperimental inquiry was made to ascertain whether *S* understood the instructions and whether, for any reason, an attempt was made to learn the incidental items. Each *S* received two learning scores, one being the number of correctly reproduced intentional items and the other being the number of correctly reproduced incidental items.

Experimental design.—Twelve groups of 16 *Ss* each were run in a standard $2 \times 2 \times 3$ factorial design. The three independent variables were meaningfulness of the intentional items, meaningfulness of the incidental items, and number of presentations. The design was balanced for top and bottom order of lists, and for the position of the intentional items. With an overall total of six possible list combinations, two possible sequences of the items from the two lists within the pairs (i.e., A-B or B-A), two possible positions for the intentional items, and three degrees of practice, a total of 72 task variations were required for the balanced design.

Subjects.—The *Ss* were students in lower division psychology courses at the University of California. They did not know the purpose of the experiment and were assigned to the conditions by rotation. They served in groups, averaging 13 in number. Each of these groups included several different experimental conditions for which the instructions were identical but which differed with regard to the stimulus lists given to *Ss*. Of necessity, two-presentation, four-presentation, and eight-presentation conditions had to be run separately. Otherwise, the experimental conditions were assigned by rotation to the members of each laboratory group with correction for equal *Ns*. The two-presentation, four-presentation, and eight-presentation treatments were assigned to the groups by rotation.

A total of 237 *Ss* were tested in order to obtain the 192 required by the experimental design. Of these, 22 indicated upon inquiry that they did not understand the instructions or that they made some attempt to learn the incidental items. They were discarded and replaced. Of these, 7 were in the two-presentation group, 8 in the four-presentation group, and 7 in the eight-presentation group. In filling out the laboratory groups during the final phase of the experiment, 23 extra *Ss* were run. Where there were extra *Ss* for any of the 72 task variations, the *Ss* to be used in the analysis were selected randomly.

RESULTS AND DISCUSSION

Incidental learning.—The incidental learning scores were analyzed in terms of the $2 \times 2 \times 3$ factorial design. The mean numbers of correctly reproduced incidental items are given in Table 2. For example, the first entry indicates that an average of 2.31 high-M items were learned incidentally, where the concurrent intentional items were also high-M, and there were two presentations. Table 3 summarizes the analysis of variance for the incidental learning data presented in Table 2. Because Bartlett's test showed significant heterogeneity of variance, the Freeman-Tukey square-root transformation ($\sqrt{x} + \sqrt{x+1}$) was used in this

TABLE 3

ANALYSIS OF VARIANCE OF INCIDENTAL SCORES

Source	*df*	*MS*	*F*
Meaningfulness (incident. items) (M_1)	1	26.85	21.46*
Meaningfulness (intent. items) (M_2)	1	.00	.00
No. of presentations (P)	2	35.25	28.17*
$M_1 \times M_2$	1	.04	.03
$M_1 \times P$	2	.29	.23
$M_2 \times P$	2	.53	.42
$M_1 \times M_2 \times P$	2	.15	.12
Error	180	(1.25)	—

* $P < .001$.

analysis, as well as in the analysis of intentional scores below. In both cases, Bartlett's test indicated that there was no significant heterogeneity in the transformed scores. It can be seen that meaningfulness of the incidental items, and number of presentations are highly significant sources of variance. The meaningfulness of the concurrently learned intentional items is not a significant source of variance in incidental learning. None of the interactions among the independent variables are significant.

If task competition is an important variable in Type II incidental learning, it might be expected that incidental learning would vary as a function of intentional-task difficulty. Therefore the question may be raised as to why incidental learning did not vary with the meaningfulness of the intentional items. The failure to find a relationship between intentional-M and incidental learning is surprising in view of the nonverbal studies which indicate that Type II incidental learning varies significantly as a function of intentional task variables (Bahrick, 1954; Bahrick et al., 1952). The considerably greater effects obtained in the nonverbal experiments become explicable when we consider the differences between the verbal and nonverbal situations. These situations differed greatly with regard to the activities required by the orienting tasks. These activities may facilitate or inhibit the performance of those differential responses which are assumed to be necessary for learning (Postman & Adams, 1956). Pronunciation of verbal items probably makes for very adequate differential responses as far as learning is concerned. In the verbal situation, the orienting task required that *S* make definite pronouncing responses to both the intentional and incidental items. Because of the nature of the orienting task, task competition could not prevent *S* from producing differential responses to the incidental items. At most, task competition could only reduce the frequency and intensity of *S*'s responses. In the case of the nonverbal experiments cited above, the situation was quite different. There was no specific orienting task which required *S* to respond differentially to the incidental stimuli. It was possible for *S* to completely ignore these stimuli while still conforming to the instructions of the experimenter. For example, *S* should be able to learn geometric forms intentionally without necessarily responding to their incidental colors. Therefore, large amounts of competition from the intentional task could result in failure to perform differential responses to the incidental materials.

It is evident from the foregoing analysis that the orienting task can be an important factor in determining the effects of task competition on incidental learning. Where the orienting task merely requires stimulus exposure of the incidental materials, the role of task competition can be great. But where the orienting task requires differential responding to the incidental materials, the role of task competition may be severely limited. In view of these considerations, it is not surprising that Bahrick's manipulations of the intentional task produced greater effects on incidental learning than did ours. We should predict that the effects of incentive upon Type II incidental learning would be sharply reduced if the orienting task required differential responses to the incidental items. Similarly, the effects of intentional-M on incidental learning should be increased by use of a less facilitating orienting task. For example, *S* might be required to assign "guessed" numbers to each incidental item instead of pronouncing it. The degree of task competition might then become an important factor in determining whether or not *S* makes appropriate differential responses to the incidental items.

Intentional learning.—The design of the analysis of intentional scores was identical to the design for the incidental scores. The means for

TABLE 4

AVERAGE NUMBER OF INTENTIONAL ITEMS RECALLED BY THE VARIOUS GROUPS

Presentations	Intentional Learning Items			
	High Meaningful		Low Meaningful	
	High Inc.	Low Inc.	High Inc.	Low Inc.
2	5.69	5.19	4.12	4.50
4	7.94	7.25	7.25	6.81
8	8.81	9.19	9.00	8.62

each condition are given in Table 4. The only significant sources of variance are meaningfulness of the intentional items ($F = 4.72$, $df = 1/180$, $P < .05$) and number of presentations ($F = 49.07$, $df = 2/180$, $P < .001$).

Differences between intentional and incidental learning.—Postman et al. (1955) and Postman and Phillips (1961) have found that the superiority of intentional over incidental learning decreases as the meaningfulness of the items increases. In the Type II situation, each *S* is exposed to both intentional and incidental lists. It is of interest to ascertain whether the difference between intentional and incidental scores for each *S* will also decrease as the meaningfulness of the items increases. Such a finding would be consistent with the Type I studies which used different groups of *S*s for intentional and incidental learning. In order to test this hypothesis, differences between intentional items recalled and incidental items recalled were obtained for each *S*. These were then divided by the *S*'s total learning in order to obtain relative difference scores for each *S* $\left(\frac{\text{Int.} - \text{Inc.}}{\text{Int.} + \text{Inc.}}\right)$. This resulted in measures of the difference between intentional and incidental learning which were comparable for different levels of overall learning. Each *S* in the experiment received one of four possible list combinations. These combinations can be ordered in terms of predicted average differences between intentional and incidental learning. Such an ordering involves two assumptions. First, that increased meaningfulness increases learning. This assumption has already been borne out by the significant *F* ratios for meaningfulness in the separate analyses of incidental and intentional learning. The second assumption is that the difference between intentional and incidental learning will decrease as the stimulus items increase in meaningfulness. On the basis of these two assumptions, the intentional-incidental differences for the four list combinations should be in the following increasing order: (1) Low Intentional-High Incidental; (2) High Intentional-High Incidental; (3) Low Intentional-Low Incidental; and (4)

TABLE 5

AVERAGE RELATIVE DIFFERENCES BETWEEN INTENTIONAL AND INCIDENTAL LEARNING FOR THE VARIOUS LIST COMBINATIONS

	List Combinations			
	(1) Low Int. High Inc.	(2) High Int. High Inc.	(3) Low Int. Low Inc.	(4) High Int. Low Inc.
Two Trials	19.23	43.94	60.04	69.58
Four Trials	34.92	42.10	45.73	51.46
Eight Trials	31.41	28.35	47.66	41.55
All Trials (two, four, & eight combined)	28.52	38.13	51.14	54.20

Note.—Means are based upon 16 *S*s for each list combination at each level of practice. The score for each *S* is: $\frac{\text{Int.} - \text{Inc.}}{\text{Int.} + \text{Inc.}} \times 100$.

High Intentional-Low Incidental. A significant gap between the second and third list combinations would indicate that the difference between the two kinds of learning varies as a function of the meaningfulness of the items.

Table 5 gives the mean relative differences for each list combination at each level of practice. The bottom row of the table gives the average differences when the three levels of practice are combined. It can be seen from this row that the differences increase in the predicted order. The largest difference is between List Combinations 2 and 3.

The data for two and four trials are also in the predicted order without any reversals. At eight trials, there are small reversals between Combinations 1 and 2 and between Combinations 3 and 4. However, Combination 3 shows a larger difference than Combination 2 at all levels of practice.

The difference scores were subjected to an analysis of variance with list combinations and number of presentations as the independent variables. Because the scores in this analysis are percentages, the arc-sine transformation (Angle = Arc-sine $\sqrt{\text{Percentage}}$) was used. As hypothesized, list combinations is a highly significant source of variance ($F = 6.58$, $df = 3/180$, $P < .001$). The number of presentations does not contribute significant variance to the percentage-difference scores ($F = 1.78$, $df = 2/180$). The interaction between list combinations and presentations is also not significant ($F = 1.64$, $df = 6/180$).

In order to analyze further the significant effects of list combinations, the gap test of Tukey (1949) was utilized. The overall means of the transformed scores were tested in order to determine whether any of the differences between adjacent means were significant. These means differ from those in the bottom row of Table 5 only in that the scores have been transformed. The means for List Combinations 1–4 are 27.86, 35.16, 46.63, and 49.34, respectively. A gap as large as 10.85 between adjacent means is significant beyond the .05 level for a two-tailed test and beyond the .025 level for a one-tailed test. In view of the predicted ordering of conditions, the one-tailed probabilities are appropriate. It may be seen that the only gap larger than 10.85 ($P < .025$) is between List Combination 2 (H-H) and List Combination 3 (L-L). This indicates that the higher the meaningfulness of the items, the smaller is the difference between intentional and incidental learning.

The gap between List Combinations 1 and 2 falls short of statistical significance ($.05 < P < .10$). Therefore the application of Tukey's test divides the means into two significantly separated pairs: (*a*) the pair of mean differences when incidental items are of high meaningfulness; (*b*) the pair of mean differences when incidental items are of low meaningfulness. Differences between intentional and incidental learning are significantly greater in the case of the latter pair.

With regard to the differential response hypothesis, the crucial comparison of mean intentional-incidental differences is between the two groups with unmixed lists (H-H vs. L-L). As indicated above, the intentional-incidental difference is significantly greater when both stimulus lists are of low meaningfulness than when both are of high meaningfulness. This is a confirmation of the hypothesis using intentional-incidental differences for individual *S*s. Previous supporting data from Type I experiments have involved comparisons of groups of intentional learners with groups of incidental learners.

The role of practice.—Another finding of this experiment was that both intentional and incidental learning increased reliably as a function of presentations. These data are in contrast to a previous study on Type I incidental learning (Postman & Adams, 1958). There it was found that intentional learning showed a significant upward trend as a function of presentations whereas incidental learning did not. These different findings are probably due to the difference between orienting tasks in the two experiments. It has been shown that differences between intentional and inci-

dental learning vary as a function of the nature of the orienting task (Postman & Adams, 1956). The earlier study of the effects of presentations required that *S* guess numbers assigned by chance to each syllable. *Ss* were told that each presentation of the series would have different number assignments. The *Ss* were faced with relatively novel nonsense syllables during the initial presentations of the series. With repeated presentations, there was no requirement for incidental *Ss* to respond to the syllables differentially. Therefore it might be expected that practice would have only marginal effects on incidental learning while having large effects on intentional learning. The present experiment required that *S* pronounce the items to himself in order to rate phonetic similarity. Because this form of response to the incidental items was required of *S* on every presentation, it is not surprising that incidental learning increased with practice in very much the same manner as did intentional learning.

Summary

This experiment studied "task competition" by manipulating the difficulty of both the intentional and incidental materials in a Type II incidental learning situation, in which *S* is exposed to two sets of materials, instructed to learn only one of the sets, and is later tested for the materials which he was not instructed to learn.

Twelve groups of 16 *Ss* each were run in a standard $2 \times 2 \times 3$ factorial design. The three independent variables were meaningfulness of the intentional items (high vs. low), meaningfulness of the incidental items (high vs. low), and number of presentations (two, four, eight). Scores for both incidental and intentional learning were obtained from each *S*.

The first analysis of the data studied incidental learning scores in terms of the $2 \times 2 \times 3$ factorial design described above. The meaningfulness of the incidental items was a significant source of variance in incidental learning. The meaningfulness of the intentional items did not contribute significant variance either by itself or in interaction with the other independent variables. In contrast with an earlier study, it was found that incidental learning increased significantly as a function of practice.

The second analysis of the data involved the difference scores between intentional and incidental learning for each *S*. It was predicted that differences would be greater when the items were of low meaningfulness than when they were of high meaningfulness. Analogous results have been found in the Type I situation and have been explained in terms of a differential response hypothesis. The predicted results in this Type II experiment support such an hypothesis.

References

Bahrick, H. P. Incidental learning under two incentive conditions. *J. exp. Psychol.*, 1954, **47**, 170–172.

Bahrick, H. P., Fitts, P. M., & Rankin, R. E. Effect of incentives upon reactions to peripheral stimuli. *J. exp. Psychol.*, 1952, **44**, 400–406.

Postman, L., & Adams, P. A. Studies in incidental learning: IV. The interaction of orienting tasks and stimulus materials. *J. exp. Psychol.*, 1956, **51**, 329–333.

Postman, L., & Adams, P. A. Studies in incidental learning: VII. Effects of frequency of exercise and length of list. *J. exp. Psychol.*, 1958, **56**, 86–94.

Postman, L., Adams, P. A., & Phillips, L. W. Studies in incidental learning: II. The effects of association value and of the method of testing. *J. exp. Psychol.*, 1955, **49**, 1–10.

Postman, L., & Phillips, L. W. Studies in incidental learning: IX. A comparison of the methods of successive and single recalls. *J. exp. Psychol.*, 1961, **61**, 236–241.

Thorndike, E. L., & Lorge, I. *A teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Tukey, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.

Underwood, B. J., & Postman, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received May 2, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 601–609

CONSISTENCY AND COMPLEXITY OF RESPONSE SEQUENCES AS A FUNCTION OF SCHEDULES OF NONCONTINGENT REWARD [1]

JOHN C. WRIGHT [2]

Stanford University

It is frequently noticed in experiments on learning that certain nonessential aspects of a response, or even wholly nonessential responses, are reinforced along with those responses or aspects of response upon which reinforcement is contingent (Guthrie & Horton, 1946; Morse & Skinner, 1957; Skinner, 1948).

Also, when analysis is made of orderly response tendencies as a function of random, noncontingent reinforcement of such tendencies, it is found that *S*s respond in an orderly fashion, though the patterns generating such orderliness vary greatly both within and between *S*s.[3] For example, in an arithmetic concept formation task, Jenkins (1959) found that the higher the probability of noncontingent reinforcement, the more individual *S*s tended to converge upon a single operation or rule for generating their responses. He noted that rules, rather than responses, seemed to be established by this process. Since the operations "learned" were in fact superstitious and bore no contingent relationship to reinforcement, it was impossible to predict what operation a particular *S* would converge upon, but the degree of convergence appeared to be regularly controlled by the probability of noncontingent reinforcement selected by *E*.

The present experiment is an investigation of the effect of the probability of noncontingent reward on the development of response sequences in human *S*s.[4] The purpose is to determine how much of the process of developing and testing hypotheses is determined simply by the occurrence of reward, regardless of what response aspects happen to precede reward.

METHOD

Apparatus

The *S*s were seated before a plain, black, sloping panel. Centered on the panel at chest level was a circle of 16 regularly spaced, black pushbuttons. A seventeenth button was mounted in the center of the circle. All buttons were unmarked and identical except for location on the panel. The buttons were 1.8 cm. in diameter, and the diameter of the circular array was approximately 20 cm. Above the sloping panel was a vertical panel containing a three-digit, cumulative counter visible to *S*.

There were two independent circuits within the apparatus. The recording circuit

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree in the Department of Psychology of Stanford University. The author wishes to acknowledge the advice and support of Alex Bavelas in the design and conduct of this experiment.

[2] Now at the University of Minnesota.

[3] A. Bavelas, personal communication, 1958.

[4] The term, *reward*, is used in place of reinforcement in order to distinguish the present situation from that in the Humphreys-Estes guessing situation (Estes & Straughan, 1954). In the guessing situation, *S* is instructed to *anticipate* which of two events will occur, while in the present situation *S* is to *control* which of two events occurs, and to maximize the occurrence of one of them. In the guessing situation every outcome is said to be a reinforcing event; in the present situation rewards are reinforcing, and nonrewards are not.

consisted of an electric typewriter, concealed from *S*, with 16 of its keys arranged so that each could be activated by a different button. Thus a code letter was recorded on paper each time *S* pressed one of the 16 buttons on the perimeter of the circle. The typewriter was wired so that the platen rolled up one line and the carriage returned automatically following each block of 25 recorded responses. The *S*s operated the buttons continuously, generating a typed record of the response sequence, with 25 characters in each line. On responses followed by reward, the space bar on the typewriter was activated, thereby yielding a record of the specific responses that happened to be rewarded, as well as a confirmation that the preprogramed sequence of rewarded and unrewarded trials was delivered as scheduled.

The event-scheduling circuit consisted of a 250-position stepping switch, wired to a control panel so that any of the 250 consecutive positions on the stepping switch could be activated by a single wire on the panel. Thus, a consecutive string of 250 trials could be preprogramed as to their outcomes. On each trial *S* was instructed to press one of the 16 perimeter buttons and then to press the center (seventeenth) button to "test" whether or not his preceding response had earned him a point. In fact, the center test button merely advanced the stepping switch to the next position, where, if the corresponding hole in the control panel had been wired, *S* received a reward, consisting of an additional point on the counter and a buzzing sound for the duration of the test response. On trials when the panel had not been wired for a reward, the depression of the test button produced only the click of the advancing stepping switch and no additional points on the counter.

TABLE 1

SCHEDULES OF NONCONTINGENT REWARD FOR THE FIVE GROUPS

Group	Block Number	Value of π	Number of Rewarded Trials per Block
20	1 to 10	.20	5
50	1 to 10	.50	12 on odd numbered blocks 13 on even numbered blocks
80	1 to 10	.80	20
Rising	1	.04	1
	2	.08	2
	3	.20	5
	4	.32	8
	5	.44	11
	6	.56	14
	7	.68	17
	8	.80	20
	9	.92	23
	10	.96	24
Falling	Same as rising, but in reverse order		

Procedure

The *S*s were recruited from the introductory psychology course at Stanford University, to take part in "an experiment in problem solving." Eighty *S*s were randomly assigned to five groups. The sex distribution was ignored. The *S*s were instructed to operate the buttons, one at a time, so as to maximize the number of buzzes, indicated by the counter. They were told that one or more correct selections, patterns, combinations, or sequences of buttons would be rewarded. Each response was to be tested by pressing the center button before the next response was made. The *S*s were told to work at a steady, comfortable pace, since the duration of the experiment was determined by the number of responses rather than elapsed time.

All *S*s then operated continuously and without comment or interruption by *E* for 325 trials, taking 30–40 min. to do so. The trials were divided into 13 blocks of 25 trials each. During Blocks 11 and 12, no rewards were given, and in Block 13 every response was rewarded. These two conditions in the last three blocks were the same for all groups, and will henceforth be referred to as "extinction" and "recovery," respectively. It was felt that 0% and 100% reward were special conditions that might have discontinuous effects on response sequences, and accordingly they were applied only after all the other conditions in order to obtain rough estimates of behavior at these extreme values.

Blocks 1 to 10, the first 250 trials, contained the experimental variations. In the following description π is defined as the probability of reward, and $\pi = .20$, say, means that within the block in question exactly 20% of the trials were rewarded, regardless of *S*'s responses, the sequence being randomized once for all *S*s in that group on that block. Whenever the same π value was called for in a different group of *S*s, *or* in a different block for the same group, the sequence of rewarded and unrewarded trials

was again randomized. Five groups of 16 *Ss* each were run under different schedules of noncontingent reward prior to extinction and recovery, as shown in Table 1. The first three groups will be referred to as the "fixed groups" because for them π was constant throughout the first 10 blocks. The other two groups will be referred to as the "sliding groups" because the value of π changed in each successive block.

The purpose of the sliding group schedules was to determine whether regular changes in π would produce any special effects beyond those due to the various π values themselves. Only nonsignificant suggestions of such effects were obtained, and therefore only effects due to π values, regardless of order of presentation, will be considered.

Scoring the Results

The nature of the experiment required that an entirely new method of scoring be developed, one that would be equally sensitive to orderliness in responding, regardless of what peculiar and complex hypothesis an individual *S* used to generate a response sequence. One cannot assume that the scorer's ingenuity is always equal to or greater than that of *S*. Nor can one get uniformly intelligible verbal descriptions of rules for generating response sequences from *Ss* themselves. Consequently, a scoring system was devised to measure any response sequence on a dimension from orderliness and predictability to randomness and unpredictability. The measure finally obtained for various transformations of the scores is called *response uncertainty* and is measured in runs per block. With the restrictions noted below, it is a rough measure of randomness in response sequences. It is inversely analogous to the "transmitted information" measure of Shannon (1948) and to the "stereotypy" measure of Miller and Frick (1949).

Zero-order scoring.—The raw data were in the form of continuous response chains, the responses being selected by *S* from among the 16 response buttons. The buttons were arbitrarily numbered clockwise from 1 to 16, and each response was recorded as the number of the button pressed. The *Ss* had no knowledge of this arbitrary notation. In each block of 25 consecutive responses, the zero-order score was defined as the number of runs of a single button in each block. A minimum score of 1 run would indicate that *S* pressed a single button 25 times in succession in that block, and was maximally consistent. A maximum score of 25 runs in a block of 25 responses would indicate that *S* never pressed the same button twice in succession and therefore was minimally consistent in that block. The number of runs at this level of scoring is called zero-order response uncertainty. It measures inconsistency of responding, with the restriction that consistency, at the zero-order level of scoring, can result only from a tendency to repeat responses, or to eliminate responses from consideration. It is assumed that at this level restriction of the range of responses and the tendency toward runs on a single response are part of the same process of convergence.

First-order scoring.—Most *Ss* converged on response rules rather than on specific responses. The most common example is constant-interval rotation. In order to detect the consistent use of such rules, the response chain of each *S* was transformed as follows: Each response was rescored by counting clockwise the number of steps (buttons in the array) required to move from that response to the next. The first-order transformation for each response indicates how far around the circle of buttons *S* "moved" from each response to the next in succession. Each block of 25 responses yielded 24 such intervals. The twenty-fifth interval for each block was defined as the interval between the last response on that block and the first response on the next block. Using the transformed data, first-order response uncertainty was defined as the number of runs of a single interbutton interval in one block. A score of 1 would indicate that the distance between successive responses was constant, and therefore that *S* was maximally consistent. Conversely, a score of 25 would indicate that *S* never used the same interbutton interval twice in succession, and was therefore minimally consistent. A pattern of consistent button repetition (low zero-order response uncertainty) would also appear as consistent in first-order scoring, since for button repetition, every interbutton interval is zero.

Second-order scoring.—A few *Ss* converged upon response rules other than repetition or constant-interval rotation, so that consistency would not be detected by either zero- or first-order scoring. Therefore, the intervals produced by first-order scoring were again transformed. The 16 intervals were reduced to 9 by taking the shortest interval between responses, regardless of rotational direction. The absolute difference between each interbutton interval and the immediately following interbutton interval yields a transformed sequence of 25 scores per block. The

TABLE 2

ILLUSTRATION OF SCORING TRANSFORMATIONS AND MEASURES OF RESPONSE UNCERTAINTY FOR A HYPOTHETICAL *S*

Order of Scoring	Scoring Transformations / ←——— one block of 25 trials ———→ / ← next block	Response Uncertainty[a]
Zero	/ 2 2 2 2 2 2 3 4 5 6 7 8 9 11 13 15 1 3 4 6 9 13 2 8 / 15 5 10	19 runs
First	/ 0 0 0 0 0 1 1 1 1 1 1 1 2 2 2 2 2 1 2 3 4 5 6 7 / 6 5	10 runs
Second	/ 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 1 1 1 1 1 1 1 / 1	6 runs

Note.—Rewards are not indicated in the table.
[a] In runs. Each run is underlined.

number of runs of these scores per block defines second-order response uncertainty. Again the range is from 1 (maximum consistency) to 25 (maximum inconsistency). Second-order scoring, like first-order scoring, is sensitive to response repetition and to response rules of constant interval rotation, but it can also detect consistent use of progressively changing interval rotation, single and double button alternation, and single interval alternation as response rules. Table 2 illustrates the scoring method and the response uncertainty values obtained at each level of scoring for a hypothetical *S* on a single block of 25 responses.

It can be shown that still higher orders of scoring would detect still more subtle and complex kinds of consistency in responding, but inspection of second-order transformations for all *S*s revealed that only 2 of the 80 *S*s made use of response patterns which, though apparent in their second-order scores, would require third-order scoring in order to be reduced to runs.

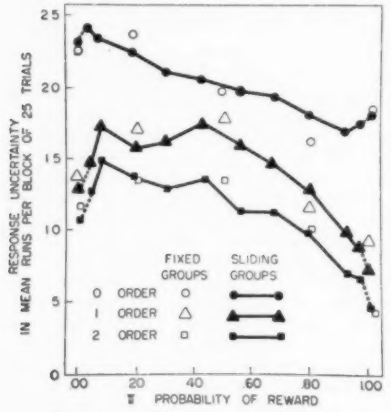

FIG. 1. Response uncertainty as a function of π.

Null values for each order of response uncertainty have been calculated for independent, random sequences of responses. Given the response on Trial $n - 1$, the probability of a response repetition on Trial n is simply 1/16. The expected number of repetitions per block is thus 25/16, or 1.56. The expected value of zero-order response uncertainty for a single *S* in a single block is thus 23.44. The same expected value holds for first-order response uncertainty, since given any of 16 possible intervals between Response $n - 1$ and Response n, then only 1 of the 16 responses possible on Trial $n + 1$ will generate a repetition of that interval, thus reducing the number of runs in that block by one.

The null value for second- and higher-order response uncertainty is somewhat lower, because direction of rotation is ignored, yielding 9 rather than 16 possible intervals. Two of the 9 intervals (0 and 8) can be generated by only one response: 0 by a button repetition and 8 by selection of the diametrically opposed button. Each of the remaining 7 intervals can be generated by either of two responses, depending on the direction of rotation. The probability of a repeated interval change is thus $(2/16 \times 1/16) + (14/16 \times 2/16)$, or .117. The expected number of repeated scores per block is thus 2.93, and the expected value of second-order response uncertainty is 22.07, for a single *S* on a single block of trials.

Complexity scoring.—By subtracting second-order response uncertainty from first-order response uncertainty, it is possible to measure to degree to which *S*s used consistent response rules more complex than either repetition or constant interval rotation. The resulting score is called *complexity*, defined as that component of general consistency due ex-

clusively to the use of the more complex patterns detectable only at the second-order scoring level. Under conditions of random response selection, the expected value for complexity is 1.37, the difference between the null values for first- and second-order response uncertainty.

Results

Figure 1 shows response uncertainty (mean number of runs per block) as a function of π (probability of noncontingent reward) for zero-, first-, and second-order scores. Response uncertainty for the π values of 0 and 1 have been estimated from the extinction and recovery blocks, respectively, and are omitted from the statistical analyses.[5] Since the order of presentation of π values for the rising and falling groups is counterbalanced, and since response uncertainty for the two groups did not differ significantly, either overall or in interaction with π values, their scores have been combined in a single curve labeled "sliding groups."

Zero-Order Scores

There is an approximately linear, negative relationship between zero-order response uncertainty and probability of reward. Button repetition is greatest when π is highest. Response uncertainty for the fixed groups is matched by that for the sliding groups at corresponding π values. To test the effect of π on response uncertainty, analysis of variance was supplemented with trend analysis using orthogonal polynomials (Lewis, 1960). For the three fixed groups, the linear component of the between-groups variance due to π values is significant at the .005 level, and the residual is nonsignificant. For the rising and falling groups combined, only the linear component of variance due to π values is significant ($P < .001$). No test of the interaction between *S*s and π values is possible, and the interaction between π value and condition (rising vs. falling) is not significant. It is concluded that the tendency to repeat a button is a linear positive function of the probability of noncontingent reward.

First- and Second-Order Scores

Since successively higher orders of scoring take into account convergence (reduction in response uncertainty) due to the consistent use of progression rules, as well as that due to button repetition, it is evident that higher orders of scoring cannot yield higher response uncertainty, and will usually yield lower response uncertainty at all π values. This result is in part a product of the scoring scheme and is clearly indicated in Fig. 1, where higher order scoring yields lower response uncertainty.

The most striking finding is that the shift from zero-order scoring (for response repetition) to second-order scoring (for all kinds of response orderliness detected) produces a marked change in the shape of the π function. The function becomes nonmonotonic, having a flat maximum between .08 and .50. Below .50 response uncertainty does not rise, and as π approaches zero, uncertainty shows some indication of declining. Since the curve for second-order scores roughly parallels that for first-order scores and is a more general measure of convergence, only the analysis of second-order scores will

[5] Combined means for the fixed groups (20, 50, and 80) and combined means for the sliding groups (rising and falling) are used as estimates of two sets of extinction and recovery points. No significant differences among groups were found for the extinction or the recovery values of response uncertainty.

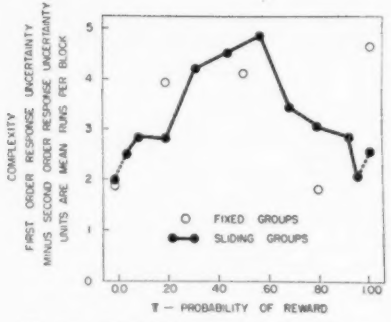

FIG. 2. Complexity of response rules as a function of π.

be reported. For the three fixed groups, the linear component of the between-groups variance is significant at the .025 level, and the residual is not significant. Therefore, on the basis of the three fixed groups alone, we may not conclude that the relationship is curvilinear. Doubtless significant curvilinearity would appear if the extinction and recovery estimates could legitimately be included in the analysis, to yield five data points instead of three. For the sliding groups, which again correspond rather closely to the fixed groups at the appropriate π values, the linear component of variance due to π values is significant at the .0001 level, and the quadratic component is significant at the .001 level. No other components approach significance. This means that there is a strong negative relationship between π and response uncertainty in general, and that there is also a highly significant curvilinear trend.

Complexity Scores

Figure 2 shows complexity as a function of π. As in Fig. 1, the rising and falling groups are combined. Again the extinction and recovery values are estimates and are excluded

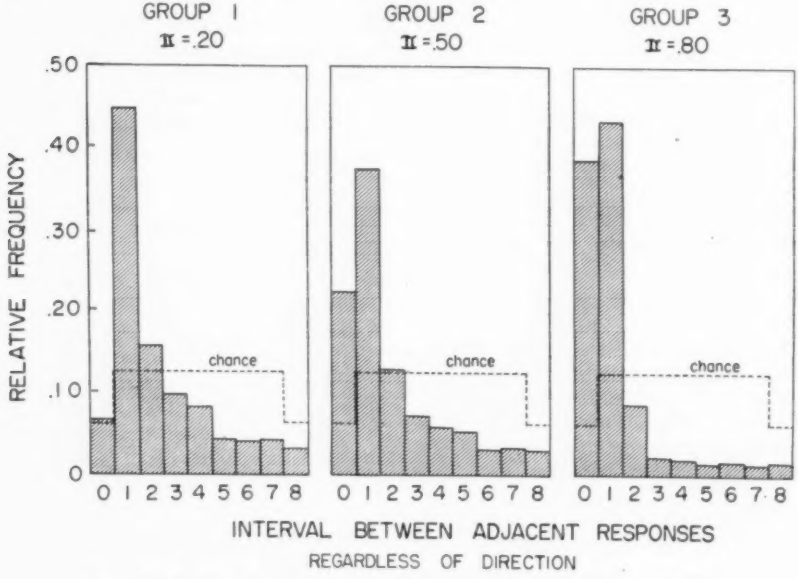

FIG. 3. Frequency distributions of response intervals, fixed groups only; N = 16 *Ss* per group, 250 responses per *S*.

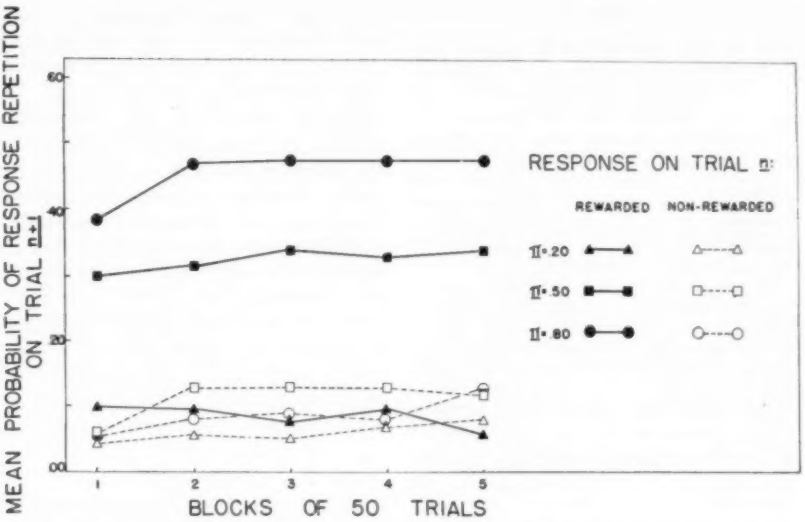

FIG. 4. Mean probability of a response repetition on Trial $n + 1$ as a function of reward and nonreward on Trial n, fixed groups only; in blocks of 50 trials.

from the analysis. Trend analysis of the complexity scores yields no significant trends due to π values for the fixed groups. For the sliding groups, however, π values contribute a quadratic trend significant at the .001 level, no significant linear trend, and no higher order trends. Therefore, there is a significant curvilinear relationship between π and complexity.

Distribution of Response Intervals

Figure 3 shows the relative frequency distributions of the intervals between successive responses for the three fixed groups, based on 4,000 trials per group. Clearly, intervals of zero (response repetition) and one (single step rotation) predominate. The only significant between-groups difference is in the relative frequency of response repetition, where the linear trend due to π values is significant at the .005 level.

Single Trial Effects

Repetition.—As π is increased, response repetition is increased. Analysis by single trials is summarized in Fig. 4, which shows the proportions of rewarded and nonrewarded trials that were followed on the next trial by response repetition, for the three constant groups over blocks of 50 trials. Not only does reward have a clearly reinforcing effect on the preceding response, but the magnitude of the effect varies positively with π. Table 3 summarizes the trend analyses for probability of response repetition. Note that both trends over trial blocks and trends over π values were tested. The reduction in zero-order response uncertainty as π increases is thus due not only to the increased number of rewards, but also to the increased effectiveness of each reward as a reinforcement for button repetition.

Rotation.—Although Fig. 3 shows

TABLE 3

ANALYSIS OF VARIANCE OF PROBABILITY OF RESPONSE REPETITION AS A FUNCTION OF π VALUES, REWARD VS. NONREWARD, AND TRIALS

Source	*df*	*MS*	*F*
π values (G)	2	1.6788	
Linear trend	(1)	3.1383	7.26**
Residual	(1)	0.2193	0.51
Ss within groups	45	0.4320	error
Total between *Ss*	47	0.4850	
Trials (T)	4	0.0265	
Linear trend	(1)	0.0675	3.67
Quadratic trend	(1)	0.0164	0.89
Residual	(2)	0.0110	0.60
T × G	8	0.0092	0.50
T × *Ss* within groups	180	0.0184	error
Reward vs. Nonreward (R)	1	4.8823	18.93***
R × G	2	1.2240	4.75*
R × *Ss* within groups	45	0.2579	error
R × T	4	0.0013	0.09
R × T × G	8	0.0105	0.75
R × T × *Ss* within groups	180	0.0140	error
Total within *Ss*	432	0.0580	
Grand total	479		

* $P < .05$.
** $P < .01$.
*** $P < .001$.

that the relative frequency of one-step rotation is not significantly affected by π, a trial-by-trial analysis was undertaken to determine the effect of reward on the probability that a one-step interval is followed by another one-step interval in the same direction of rotation. Since the trends and interactions involving π values were inconsistent and nonsignificant, only the effects of reward vs. nonreward are shown in Fig. 5. The only significant effect is that due to reward vs. nonreward ($P < .01$), and it is concluded that a nonreward is more effective than a reward in eliciting continued one-step rotation.

DISCUSSION

Given a minimum level of reward, *Ss* will converge on and repeat consistently some response or rule as a function of the relative frequency of reward pro-

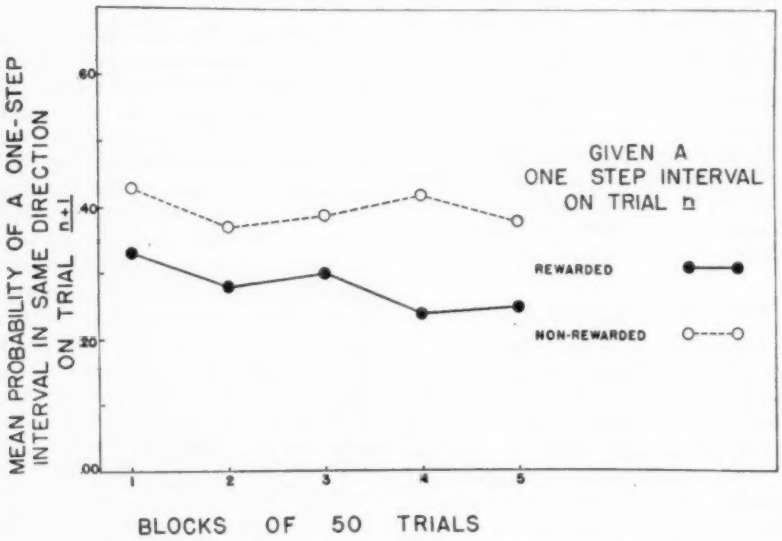

FIG. 5. Mean probability, given a one-step interval on Trial n, of a one-step interval in the same rotational direction on Trial $n + 1$ as a function of reward and nonreward on Trial n. (In the response sequence i, j, k, the interval, $i - j$ is rewarded if and only if a reward occurs after Response j and before Response k.)

vided. In the case of response repetition, the overall density of reward interacts positively with the occurrence of a single reward to enhance its reinforcing power. Regardless of the overall reward density, however, a nonreward is more effective than a reward in producing a pattern of systematic rotation. Thus it might be argued that response repetition is an indication that *S* has a "solution," and one-step rotation is an indication that *S* has a systematic scanning strategy.[6] A qualification is the fact that at high densities of reward the scanning strategy may also be the solution.

Patterns of responding more subtle than repetition or rotation of responses occur most consistently when the uncertainty of outcome is greatest (i.e., when $\pi = .50$). This finding with respect to complexity suggests that when the inconsistency of feedback precludes both satisfactory convergence on a solution and orderly pursuit of a systematic scanning strategy, then *S* is forced to try more complex patterns, or simply to guess.

Summary

Eighty *S*s were instructed that their responses on a circular array of pushbuttons determined which of two outcomes, reward or nonreward, would follow each response, but were in fact presented with preprogramed, random sequences of outcomes which differed systematically only in the relative frequency of reward. Various probabilities of reward were scheduled. Consistency of responding, defined as the repetition of any single button, was a positive, linear function of the probability of reward. The relative reinforcing power of a single reward for the next response was also a positive linear function of the probability of reward. When consistency due to orderly sequences of button selection was taken into account as well, the relation was positive, but curvilinear, consistency being greatest at high probabilities of reward, intermediate at low probabilities, and low at intermediate probabilities of reward.

Complexity of the response rules was greatest when the outcomes were most inconsistent (50% random reward), and least when the outcomes contained either high or low probabilities of reward. Complexity was also high when a period of 100% reward followed a relatively long period of nonreward.

It was concluded that superstitious response preferences and patterns are established as a function of the probability of reward, but that rotational sequences for the systematic testing of responses may also be established as orderly scanning rules, especially following nonreward. A distinction was made between collecting and using information—behaviors that are differentially influenced by the uncertainty of outcome and the probability of reward.

[6] Miller, Galanter, and Pribram (1960) present an excellent discussion of such plans for seeking information and finding solutions. Evidence concerning the distinction between organizing a search for information and organizing the use of information is offered by Pribram (1959) and by Jenkins (1959).

References

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Guthrie, E. R., & Horton, C. P. *Cats in a puzzle box.* New York: Rinehart, 1946.

Jenkins, H. M. An exploratory study of the development of concepts with noncontingent reinforcement. *Bell Lab. tech. Memo.*, 1959, No. 59-122-3.

Lewis, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

Miller, G. A., & Frick, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.

Miller, G. A., Galanter, E., & Pribram, K. H. *Plans and the structure of behavior.* New York: Holt-Dryden, 1960.

Morse, W. H., & Skinner, B. F. A second type of superstition in the pigeon. *Amer. J. Psychol.*, 1957, **70**, 308–311.

Pribram, K. H. On the neurology of thinking. *Behav. Sci.*, 1959, **4**, 265–287.

Shannon, C. E. A mathematical theory of communication. *Bell Sys. tech. J.*, 1948, **27**, 623–656.

Skinner, B. F. "Superstition" in the pigeon. *J. exp. Psychol.*, 1948, **38**, 168–172.

(Received May 5, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 610–615

SACCHARINE CONCENTRATION AND DEPRIVATION AS DETERMINANTS OF INSTRUMENTAL AND CONSUMMATORY RESPONSE STRENGTHS [1]

HARRY L. SNYDER [2]
Johns Hopkins University

The importance of the consummatory response in the reinforcement process has been suggested by several authors. Sheffield, Roby, and Campbell (1954) have indicated that running speed increases with the average rate of ingestion of various sweet solutions. Similarly, Spence (1956) has suggested that a monotonic relationship should hold between running speed and the vigor of the consummatory response. If such consummatory response theories are to have any general utility, it follows that any variable or combination of variables which affects the strength of consummatory behavior should similarly affect the strength of the preceding running response. The present experiment examines this theoretical approach by analyzing concomitant variation in running and consummatory behavior as related to both deprivation and saccharine concentration.

In addition, the data provide information about the manner in which drive and incentive combine to determine the strength of a running response.

METHOD

Subjects and Apparatus

The *Ss* were 120 experimentally naive male rats of the Sprague-Dawley strain which were obtained from Sprague-Dawley, Incorporated, Madison, Wisconsin. The *Ss* were 130 to 150 days old at the start of the experiment. They were housed in individual cages throughout the study.

The *Ss* were trained in a 68-in. Plexiglas-covered, wooden alleyway. Guillotine doors separated the start and goal boxes from the running alley. The start box was 9 in. long and 6 in. wide, and the goal box was 11 in. long and 8 in. wide. The running alley was 4 in. wide. The entire alleyway was 9 in. high and was painted a flat black.

A microswitch, activated by raising the start-box door, and two photoelectric cells in circuit with Standard Electric timers automatically recorded starting latencies and running times. The photocells were located 1 in. and 45 in., respectively, past the start-box door.

A 3½-in. square piece of Plexiglas was centered in the wall at the far end of the goal box. A drinking tube was located in a ¾ × ⅜ in. vertical slot cut in the Plexiglas. The drinking tube consisted of a piece of ¼-in. brass tubing which contained a 2-mm. fluid hole.

With each lick upon the drinking tube, a pumping system delivered .004 ml. of fluid at the tip of the drinking tube. The pumping system, which is described in more detail elsewhere (Hulse, 1960), consisted of an infusion pump operated by an electronic relay circuit. The relay circuit also operated a counter and an Esterline-Angus pen recorder, Model AW, which indicated, respectively, the number of licks made upon the drinking tube, and the temporal distribution of licks. The paper speed of the recorder was 3 in. per sec. A stepping relay also activated a Standard Electric timer when *S* made the first lick of each trial. After the last lick on each trial, *E* stopped the timer. In this manner, the total time *S* took to make 150 licks was recorded.

Experimental Design

Four levels of food and water deprivation (2, 6, 12, and 22 hr. since last feeding) and

[1] This paper is based upon a dissertation submitted in partial fulfillment of the requirements for the PhD degree. The author is indebted to Stewart H. Hulse, director of the dissertation, for his helpful advice. This research was supported in part by a National Science Foundation Grant, G8712, to Stewart H. Hulse.

[2] Now at Colgate University.

three levels of saccharine concentration (0.25, 4.50, and 10.0 gm. saccharine added to 1 liter of water) were combined factorially to produce 12 experimental groups. Ten *S*s were randomly assigned to each group.

Procedure

All experimental operations were carried out in an air-conditioned, humidity-controlled room under normal room illumination. These environmental conditions were maintained 24 hr. per day to avoid confounding of day-night cues with the times of running of the various groups. Temperature fluctuation throughout the experiment was less than 5° F.

Taming.—For 15 days, each *S* was fed a daily ration of 14 gm. dry Purina lab chow mixed with 29 ml. tap water to form a wet mash. This diet was maintained throughout the experiment. Following this diet-adaptation period, and continuing for 6 days, *S*s were handled freely and permitted to drink for 15 min. from a drinking tube containing the same concentration of saccharine that they would receive during training. Following each daily taming session, *S*s were returned to their living cages and fed at a time such that, for the following taming session, the appropriate deprivation state would be in effect. On the last day of taming, *S*s were handled but were not permitted to drink from the saccharine bottles.

Training.—Following taming, *S*s were given 35 training trials, 1 trial per day. Within each deprivation group, the order of running was randomized daily. On each trial, *S* was placed in the start box with the door lowered. After a delay of 2–3 sec., the door was raised and *S* was permitted to run to the goal box. After *S* entered the goal box, retracing was prevented by lowering the goal-box door. When the *S* had made 150 licks upon the drinking tube, it was removed to its living cage.

Extinction.—Following training, 10 extinction trials were given, 1 trial per day. All conditions were equivalent to the training conditions with the exception that no fluid was delivered through the drinking tube. All *S*s were permitted to lick upon the dry tube, however. Each *S* was removed from the apparatus 30 sec. after it entered the goal box.

Measures of Performance

Starting time (latency) was defined as the time which elapsed between *E*'s raising the start-box door and *S*'s breaking the first photocell beam. Running time was the time *S* took to run from the first to the second photocell. Running times were transformed to reciprocals to give running speeds.

Licking time was defined as the time which elapsed between the first and last licks each *S* made on each trial. The number of licks per trial (150) was divided by the licking time to give average rate of licking, in licks per second.

Esterline-Angus records were taken on Trials 33 and 34 of training. On each trial for each *S*, 4-sec. records were obtained following 10, 60, and 110 licks. Each 4-sec. record contained approximately 20–25 licks. The total time between the first and last licks was noted. From this total time, any time due to pauses between licks was subtracted. Such pauses were defined as intervals greater than twice the mean interlick interval for that record. The number of licks on the record was then divided by the remaining time (total time minus pauses) to give momentary rate of licking. Thus, while the average rate of licking is a mean value of licks per second throughout the entire session, the momentary rate of licking is a local rate at particular times within the session.

During extinction, licking behavior was measured by recording the number of licks each *S* made upon the dry tube during the 30-sec. goal confinement time.

RESULTS

Running Behavior

Training.—For the 35 training trials, starting latencies and running speeds for each *S* were averaged across seven blocks of 5 trials each. Analyses of variance were then used to assess the effects of deprivation (D), saccharine concentration (C), blocks of trials (Tr), and their interactions.

Starting latencies decreased significantly with increases in D ($F = 3.33$, $df = 3/107$, $P < .025$) and Tr ($F = 17.61$, $df = 6/642$, $P < .001$). No significant variation in latencies was obtained due to C ($P > .05$). None of the possible interactions was significant.

Running speeds increased significantly across Trials 1–35 ($F = 12.34$,

TABLE 1

SELECTED RESULTS FROM RUNNING SPEED AND LICKING RATE ANALYSES OF VARIANCE

Source	*df*	Running Speed (Trials 16–35)		Average Rate (Trials 16–35)		Momentary Rate (Trials 33–34)	
		MS	*F*	*MS*	*F*	*MS*	*F*
Concentration (C)	2	27,627.58	21.88***	129.33	47.37***	1.08	2.84
Deprivation (D)	3	3,989.94	3.16*	11.09	4.06**	1.30	3.42*
C × D	6	368.40	—	3.73	1.37	1.53	4.02**
Between *Ss* error	107[a]	1,262.77		2.73		0.38	

Note.—Within *Ss* effects, whenever important, are discussed in text.
[a] One *S* was lost during the experiment and is treated statistically as a missing case.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

$df = 6/642$, $P < .001$). This increase with Tr was greater for increases in both C and D, as indicated by C × Tr and D × Tr interactions ($F = 7.62$, $df = 12/642$, $P < .001$ and $F = 1.92$, $df = 18/642$, $P < .05$, respectively). The C × D interaction was not significant.

In order to further evaluate the effects of C and D, a separate analysis was used for Trials 16–35. The results of this analysis are presented in Table 1. For Trials 16–35, increases in either C or D led to increased running speeds, as shown in Fig. 1 and 2. In addition, the C × D interaction was not significant ($F < 1.00$), which indicates that the effects of C and D upon running speed are independent. Neither Tr nor

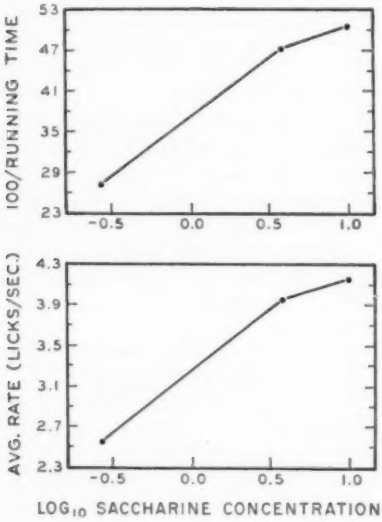

FIG. 1. The effect of saccharine concentration upon running speed and average rate of licking, Trials 16–35.

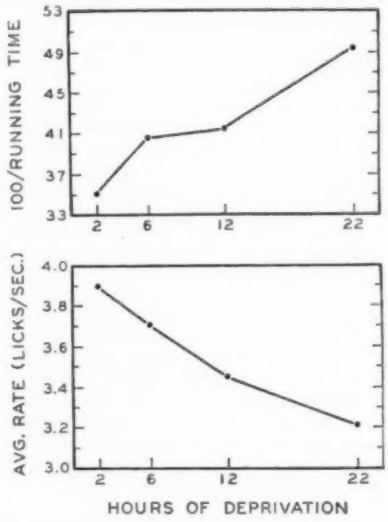

FIG. 2. The effect of deprivation upon running speed and average rate of licking, Trials 16–35.

any of the interactions involving Tr were significant for Trials 16–35.

Extinction.—Starting latencies and running speeds for the 10 extinction trials were analyzed with respect to individual trials rather than blocks of trials.

Increases in both C and D led to faster running speeds during extinction ($F=17.28$, $df=2/107$, $P<.001$, and $F = 5.72$, $df = 3/107$, $P < .01$, respectively). The interaction between C and D was not significant.

Significant C × Tr and D × Tr effects were obtained for the 10 extinction trials ($F=2.74$, $df=18/963$, and $F = 2.86$, $df = 27/963$, respectively, both Ps $< .001$). The C × Tr interaction indicates that, for the three C groups, the speed curves converge over the 10 trials. However, the D × Tr interaction presents a different picture. In this case, inspection of the group means indicates that the 22-hr. deprived *S*s show very little, if any, extinction, while the other three D groups extinguish relatively rapidly.

Consummatory Behavior

Average rate of licking.—The average rate of licking for Trials 16–35 (Fig. 1 and 2) increases with saccharine concentration ($P < .001$), but *decreases* with deprivation ($P < .01$). The C × D interaction was not significant (Table 1). Neither Trials nor any of the interactions involving Trials were significant.

Momentary rate of licking.—Momentary rates for Trials 33 and 34 varied significantly with both D and the C × D interaction, as shown in Table 1. The significant D effect was nonmonotonic and is probably due to the C × D interaction, which is illustrated in Fig. 3. Figure 3 shows that the momentary rate of licking increases as a function of saccharine concentration for *S*s under relatively low deprivation (2 and 6 hr.), remains nearly constant for *S*s under moderate deprivation (12 hr.), and decreases as a function of concentration for highly-deprived *S*s (22 hr.). For convenience, the 2- and 6-hr. groups have been combined in Fig. 3 since they did not differ significantly.

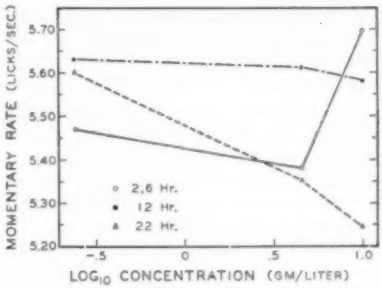

FIG. 3. Momentary rates of licking for combinations of deprivation and saccharine concentration, Trials 33 and 34.

Momentary rates of licking also decreased as licking within a session progressed ($F = 1010.67$, $df = 2/211$, $P < .001$). Moreover, this decrease with number of licks was greater for higher concentrations as indicated by a small but highly significant C × Number-of-Licks interaction ($F = 3.67$, $df = 4/211$, $P < .01$).

Extinction of licking behavior.—The number of licks made upon the dry tube during the 10 extinction trials did not vary significantly with either C, D, or C × D.

DISCUSSION

Running behavior.—The present results agree with previous experiments (Reynolds & Pavlik, 1960; Weiss, 1960) which have shown that deprivation and incentive have independent effects upon strength of response. The results thereby add to the evidence supporting Spence's (1956) notion that the theo-

retical relationship between D, K, and response strength is defined by the formula: ${}_sE_R = {}_sH_R \times (D + K)$. This idea is opposed to Hull's (1951) theory which suggests a statistical interaction between D and K, according to the formula: ${}_sE_R = {}_sH_R \times D \times K$.

Average rate of licking.—A consummatory response theory of reinforcement is supported quite well insofar as the effects of saccharine concentration in this experiment are concerned. Increases in C produce increases in both running speed and average rate of ingestion (see Fig. 1). However, increases in deprivation produce increases in running speed but *decreases* in average rate of ingestion (see Fig. 2). This decrease is not affected by saccharine concentration as indicated by the nonsignificant C × D interaction. Hence, the utility of a consummatory response theory of reinforcement may well be limited to variables dealing with the nature of the reinforcer, per se, such as sweetness or volume, but apparently does not extend to variables dealing with the deprivation state of the organism. Quite obviously, these results place a severe restriction upon the general usefulness of such theories.

Momentary rate of licking.—When they lick, rats lick faster for higher concentrations if they are slightly deprived, but lick faster for lower concentrations if they are greatly deprived (see Fig. 3). It would appear, as a first approximation, that this interaction is related to sensory events. Specifically, it may be the case that the effect of deprivation is to change the threshold for sweet stimulation and thus cause the taste receptor stimulation for saccharine solutions to vary complexly with the deprivation state of the rat. Indeed, there are data to indicate that both threshold and preference for sweet solutions are altered by deprivation (Bacon, Snyder, & Hulse, 1962; Campbell, 1958).

In a related manner, licking behavior within a session may also be determined by sensory events. The momentary rate decreases as a function of the number of licks, and this decrease with number of licks is greater for both higher concentrations and greater volumes per lick (Snyder & Hulse, 1961). Thus, when the rate of ingestion is directly contingent upon an instrumental rate of responding, there is some regulatory mechanism for ingestion which reduces this rate of responding, and the reduction in rate is greater for larger volumes or higher concentrations. Whatever the locus of the mechanism, it operates both for licking, as shown here, and for bar pressing, as indicated by Collier and Siskel (1959).

Summary

Three levels of saccharine concentration (0.25, 4.50, and 10.0 gm/l) and four levels of food and water deprivation (2, 6, 12, and 22 hr.) were combined factorially to produce 12 experimental groups. Ten male albino rats per group were given 35 training and 10 extinction trials, 1 trial per day, in a straight alleyway. On each trial, *Ss* were permitted to make 150 licks upon a drinking tube which supplied the saccharine solution. Running speed and consummatory behavior were measured.

The results were: (*a*) Running speed increased with both deprivation (D) and concentration (C). The C × D interaction was not significant. (*b*) The average rate of licking (licks/sec measured over the entire reinforcement session in the goal box) increased with C but *decreased* with D. (*c*) The momentary rate of licking (local rate within bursts of licking) varied complexly with the C × D interaction. Slightly-deprived rats lick faster for higher saccharine concentrations, but greatly-deprived rats lick faster for lower concentrations.

The additive relationship between C and D for running speeds was interpreted as supporting Spence's theoretical formulation, ${}_sE_R = {}_sH_R \times (D + K)$. Although the monotonic relationship between running speed and average rate of licking for the C variable firmly supports a consummatory response theory of reinforcement, the negative relationship between these response measures for the D variable clearly indicates a severe limitation to such a theory.

References

Bacon, W. E., Snyder, H. L., & Hulse, S. H. Saccharine preference in satiated and

deprived rats. *J. comp. physiol. Psychol.*, 1962, **55,** 112–114.

CAMPBELL, B. A. Absolute and relative sucrose preference thresholds for hungry and satiated rats. *J. comp. physiol. Psychol.*, 1958, **51,** 795–800.

COLLIER, G., & SISKEL, M., JR. Performance as a joint function of amount of reinforcement and inter-reinforcement interval. *J. exp. Psychol.*, 1959, **57,** 115–120.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1951.

HULSE, S. H., JR. A precision liquid feeding system controlled by licking behavior. *J. exp. Anal. Behav.*, 1960, **3,** 1–3.

REYNOLDS, W. F., & PAVLIK, W. B. Running speed as a function of deprivation period and reward magnitude. *J. comp. physiol. Psychol.*, 1960, **53,** 615–618.

SHEFFIELD, F. D., ROBY, T. B., & CAMPBELL, B. A. Drive reduction versus consummatory behavior as determinants of reinforcement. *J. comp. physiol. Psychol.*, 1954, **47,** 349–354.

SNYDER, H. L., & HULSE, S. H. Effects of volume of reinforcement and number of consummatory responses on licking and running behavior. *J. exp. Psychol.*, 1961, **61,** 474–479.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

WEISS, R. F. Deprivation and reward magnitude effects on speed throughout the goal gradient. *J. exp. Psychol.*, 1960, **60,** 384–390.

(Received May 10, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 616–620

CONCEPT IDENTIFICATION AS A FUNCTION OF OBVIOUSNESS OF RELEVANT AND IRRELEVANT INFORMATION [1]

E. JAMES ARCHER

University of Wisconsin

In a number of previous studies (e.g., Archer, 1954; Archer, Bourne, & Brown, 1955; Brown & Archer, 1956; Bulgarella & Archer, 1962) a binary-digit scale ($\log_2$) was used to describe the amount of information in stimuli used in concept identification tasks. In addition the information could be qualitatively described as relevant, irrelevant, or redundant. In all of the studies 1 bit of information from one dimension was treated as equivalent to an equal amount of information from any other dimension. As a first approximation, such an application of information analysis is an oversimplification. For example, the Brown and Archer (1956) study indicated that information concerning the physical location of the stimulus apparently was less obvious to *S* than information concerning form or color. Concepts having the spatial location of the stimulus relevant proved to be more difficult than others not having spatial location relevant. It seems reasonable to suppose that the obviousness of information is a manipulable variable and that such a characteristic of the information affects concept identification. Furthermore, at least one simple prediction can be made concerning the relationship between obviousness and relevance: there should be a significant interaction between the two variables, i.e., if the relevant information were also obvious, the concept should be easier to discover than if the information were less obvious. On the other hand, if the irrelevant information were obvious, the concept should be more difficult to discover than if the information were less obvious. The first part of this hypothesis is self-evident; the latter part assumes that when *S*'s attention is drawn to irrelevant information, he will attempt a solution involving information which should not be used, and when the irrelevant information is less obvious *S* is better able to ignore the information which should be ignored.

In order to assess the generality of this prediction, it seemed desirable to study two distinctly different dimensions. Because of the simplicity and accuracy of control, the dimensions of *form* and *size* were selected.

PROCEDURE

Subjects.—A total of 128 students (64 men and 64 women) from introductory psychology courses at the University of Wisconsin served individually as *S*s in this study.

Conditions.—As may be seen from Table 1 there were 16 combinations of stimulus conditions of Manipulated Dimension, Obviousness, Relevance (of Manipulated Dimension), and Level of Irrelevant Information arranged in a 2^4 orthogonal design.

Since the effect of sex of *S* was being studied, 4 men and 4 women, otherwise randomly selected, were assigned to each of the 16 combinations of stimulus conditions.

[1] This study was supported by a National Science Foundation grant (G-1691). The author is indebted to Matilda E. Levine who made the stimulus materials and gathered the data of the main study.

This report was prepared while the author was a National Science Foundation Senior Postdoctoral Fellow at the Applied Psychology Research Unit, Cambridge, England.

Apparatus.—The stimuli were back-projected onto an 8 × 10 in. opal-glass screen by a Dunning Animatic 16-mm. strip-film projector. The screen was mounted in the wall of the booth above *S*'s control console. The booth was made of sound-absorbent material 4 × 4 × 8 ft. and painted flat black inside. The control console had four equally spaced lever-action switches. Above each switch was a neon glow lamp.

Task.—The *S* was told that he would see different patterns and that his task was to assign each of the patterns to one of four categories. He was to indicate his assignment by pressing one of the switches before him. When a switch was operated, one of the four glow lamps would immediately light over the switch which should have been pressed. The *S* was urged to press one and only one switch firmly and then release it. The glow lamp remained lit for 2 sec. and then 1 sec. later the strip-film projector advanced to the next pattern.

The task was partially self-paced since *S* could take as long as he wished to decide which switch to depress, but after responding, the next pattern automatically appeared in 3 sec.

The *S* was told that all of the patterns in a particular category would have something in common and that he would continue at the task until he had made 32 consecutively correct responses.[2]

TABLE 1

DESIGN OF THE EXPERIMENT

Manipulated Dimension	Obviousness	Relevance	Stimulus Dimensions Varied[a]		
			Relevant	Irrel. (1)	Irrel. (3)
Size	High (1 in. and ¾ in.)	Relevant	size number	color	color shade horizontal
		Irrel.	color number	size	size shade horizontal
	Low (1 in. and ¼ in.)	Relevant	size number	color	color shade horizontal
		Irrel.	color number	size	size shade horizontal
Form	High (90° and 84°)	Relevant	form dot	shade	shade number horizontal
		Irrel.	shade dot	form	form number horizontal
	Low (90° and 87°)	Relevant	form dot	shade	shade number horizontal
		Irrel.	shade dot	form	form number horizontal

Note.—This design was duplicated for 64 men and 64 women students. Four *S*s of each sex served in the 16 combinations of relevant and irrelevant information.

[a] The levels of the dimensions were as follows: Size (large or small); Number (one or two identical figures); Color (red or green); Shade (light or dark); Horizontal (figure appeared on left or right of screen); Form (square or parallelogram—87° or 84° opposite acute angles); Dot (presence or absence of a ¼-in. white dot centrally located on figure).

RESULTS

Time to criterion.—Because of heterogeneity of variance the time-to-criterion scores were logarithmically transformed and the total variance of these latter scores was then analyzed. Since this was a 2^5 orthogonal design replicated four times, each main effect and interaction had 1 *df* and the error term had 96 *df*.

As expected the amount of irrelevant information was an effective variable. The mean time to criterion was 10.25 min. and 17.98 min. for 1 and 3 bits irrelevant, respectively ($F = 30.429$, $P < .001$).

There was also a difference in difficulty depending upon the manipulated dimension. The problems involving variations in size were easier (9.36 min.) than those involving variations in form (18.87 min.), $F = 31.441$, $P < .001$. None of the other main effects was significant.

As shown in Table 2 the predicted interaction of Relevance × Obviousness was highly significant. The meaning seems quite clear; if the information (and in this case we mean the manipulated information of either size or form) is obvious and relevant,

[2] The choice of the different sizes and forms was determined by a pilot study which assured 100% discriminability of the stimuli when *S*s were asked to sort for size or form. Unless the stimuli were 100% discriminable, the *S*s would not be able to solve the concept identification task.

TABLE 2

MEAN TIME TO CRITERION (MIN.) AS A FUNCTION OF OBVIOUSNESS AND RELEVANCE OF INFORMATION

Obviousness	Relevance	
	Relevant	Irrelevant
High	12.7	15.6
Low	16.5	11.7

Note.—Each mean is based on an N of 32. Because of heterogeneity of variance, the analysis was performed on log-transformed scores; $F = 7.172$, $df = 1/96$, $P < .01$.

then the criterion is achieved fairly easily. If, however, the information is equally obvious but irrelevant, then the task proves fairly difficult. Just the opposite obtains when the information is not obvious but is relevant; the task is very difficult. Finally, when the equally unobtrusive information is irrelevant, then the task is easiest of all.

Another first-order interaction which was significant was that of Relevance × Manipulated Dimension. This relationship is shown in Table 3.

Clearly the effects of the two manipulated dimensions are different. When form was relevant the task was quite difficult; and when form was irrelevant, the task became very difficult. Exactly the opposite occurred with the variation in relevance of size; when relevant the task was fairly easy and when irrelevant it became even easier. This perplexing differential effect of the two manipulated dimensions might have remained unresolved except for the results shown in Table 4.

TABLE 3

MEAN TIME TO CRITERION (MIN.) AS A FUNCTION OF MANIPULATED DIMENSION AND ITS RELEVANCE

Manipulated Dimension	Relevance	
	Relevant	Irrelevant
Size	11.9	6.8
Form	17.2	20.5

Note.—Each mean is based on an N of 32. Because of heterogeneity of variance, the analysis was performed on log-transformed scores; $F = 14.418$, $df = 1/96$, $P < .01$.

Here we see an unexpected significant Sex × Manipulated Dimension × Relevance interaction. As is evident, the two sexes do not behave differently when size was relevant or irrelevant. However, when form was relevant, men found the task fairly easy; and when form was irrelevant the men found the task very difficult. Just the opposite effect was shown by the women. When form was relevant, the women found the task quite difficult but when form was irrelevant, they found the task much easier.

TABLE 4

MEAN TIME TO CRITERION (MIN.) AS A FUNCTION OF SEX OF S, MANIPULATED DIMENSION, AND RELEVANCE OF THE DIMENSION

Sex of S	Manipulated Dimension	Relevance	
		Relevant	Irrelevant
Men	Form	12.9	24.4
	Size	12.7	8.1
Women	Form	21.6	16.6
	Size	11.1	5.5

Note.—Each mean is based on an N of 16. Because of heterogeneity of variance, the analysis was performed on log-transformed scores; $F = 4.355$, $df = 1/96$, $P < .05$.

None of the other main effects or interactions was significant.

Errors to criterion.—A different, though related ($r = .93$, $df = 126$, $P_{r=0} < .001$), measure is that of errors to criterion. Again, because of heterogeneity of variance a logarithmic transformation was applied

to these scores. An analysis of variance yielded results similar to the time analysis; the same main effects and interactons were significant except for the Sex × Manipulated Dimension × Relevance interaction which just failed of the .05 level of confidence.

Discussion

Two conclusions seem appropriate for this study. First, while an information analysis of the stimuli used in a concept identification task provides a starting point for predicting the difficulty of the task for *S*, the performance can be greatly modified by a secondary stimulus quality such as *obviousness* of information.

As shown by the present study, the optimum conditions to enable *S* to identify a concept will be when the obviousness of the relevant information is maximized and the obviousness of irrelevant information is minimized. The ultimate, of course, of this latter condition is when the difference between (or among) the levels is zero and then the irrelevant information no longer exists.

One source of variation usually described as *individual differences* may be partly due to the perceptual biases of the individual *S* which alter the obviousness of the stimulus differences. In some cases, *S* may perceive and attend to stimulus differences which are relevant and will thereby perform well. In other cases, *S*s who perceive and attend to the irrelevant information will, of course, perform poorly. As this study shows, we can vary the probability of *S* attending to the relevant and irrelevant information.

Perhaps the most intriguing aspect of the present study is the observation of the sex difference, not as a main effect, but as the Sex × Manipulated Dimension × Relevance interaction. Since it seemed unlikely that this was a biologically determined phenomenon, a perceptual, albeit post hoc, explanation was sought.

The same stimuli which had been used in the most complex form-variation condition were shown to 40 men and 40 women students and they were asked to describe what they saw. Aside from number, color, and white dot, men usually described the stimuli as squares and parallelograms or even trapezoids. Women, on the other hand, much more frequently described the stimuli as squares and nonsquares, or not squares, or even tippy squares. While recognizing the dangers of a post hoc explanation and the especial hazard of generalizing from one sample to another sample, it is hard to resist the hypothesis that the obviousness of a level within a dimension can be modified by *S*'s labeling ability. If the men in the experimental study could label the stimuli with distinctly different names, e.g., square and parallelogram or even trapezoid, then the obviousness of the form information would be enhanced (relative to not having a label for the different levels) and if the information about form were relevant, then the task would be easier than if the information were irrelevant. Furthermore, just the opposite would occur if the women in the experimental study did not have distinctly different labels, e.g., square and nonsquare; they would find the task very easy when form was irrelevant and very difficult when form was relevant.

The results suggest the way inner speech might modify problem solving behavior. It seems reasonable to predict that any verbal pretraining which successfully modifies inner speech will thereby affect performance in problem solving tasks like concept identification.

Summary

A 2^5 orthogonal design replicated four times examined the effect of obviousness of stimulus differences upon concept identification when the manipulated information was either relevant or irrelevant. It had been predicted that there would be an interaction between relevance and obviousness so that performance would be facilitated if the relevant information were obvious, and performance would be impeded if the relevant information were less obvious; and the reverse would obtain for the irrelevant information. The prediction was supported.

An unexpected result indicated that men and women perceived the two forms of squares and parallelograms differently. Men found the task easier when *form* was relevant rather than irrelevant; and the opposite obtained for women. A post hoc explanation was offered; the differential performance might have been due to the unequal availability of distinctive labels for the forms by the two sexes.

References

ARCHER, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.

BULGARELLA, R. G., & ARCHER, E. J. Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. *J. exp. Psychol.*, 1962, **63**, 254–257.

(Received May 11, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 621–626

RECIPROCAL AUGMENTATION OF GENERALIZATION AND ANXIETY[1]

SARNOFF A. MEDNICK AND CYNTHIA WILD

University of Michigan *Yale University*

As a part of a general theory of schizophrenic behavior it has been suggested that the core of the acute breakdown in schizophrenia involves an interaction between learned fear and generalization (Mednick, 1958). It is hypothesized that under certain conditions these two factors could reciprocally augment each other. The hypothesis of reciprocal augmentation suggests that if stimuli which elicit fear reactions were repeatedly presented to *S* his general level of arousal would tend to rise. As a consequence, his responsiveness to fear producing generalization stimuli would tend to be augmented. If such generalization stimuli were to continue to evoke reactions from him, his general level of arousal would be again augmented which would again produce an increase in responsiveness to fear producing generalization stimuli. The key to the uniqueness of this process is the fact that stimuli which elicit fear responses can bring about increases in level of arousal.

The process may be conceptualized in detail as follows. Let GS_1, GS_2, GS_3, and GS_4 be a series of generalization stimuli which elicit fear responses from *S*; let A be *S*'s arousal level at the inception of the series of stimuli. When GS_1 is presented, *S* will react with fear. Since learned fear contributes to arousal (Malmo, 1957), the new momentary arousal level will be $A + \Delta 1$. One important consequence of this increment in arousal level will be the raising of the height and breadth of the gradient of generalization responsiveness (Mednick & Freedman, 1960). This augmentation of generalization will be temporary, decreasing as $\Delta 1$, the fear response to GS_1, decays. However, if GS_2 is presented soon enough after GS_1 it will stimulate *S* while he is in a state of heightened arousal and will consequently elicit a relatively larger fear response than GS_1 had elicited. This response will then send the arousal level up to $A + \Delta 1 + \Delta 2$ (less some quantity ascribable to the decay of $\Delta 1$). This temporary rise in arousal level will in turn again augment generalization responsiveness. Now, if GS_3 is presented soon enough after GS_2, it will elicit a relatively larger response, add to total arousal level, augment the response to GS_4 . . . and so on.

One distinctive and testable prediction that the reciprocal augmentation hypothesis generates suggests that a learned fear response (and generalized responses) would extinguish more slowly under massed practice than distributed practice. The longer the time between trials the greater the opportunity for the fear response from the preceding trial to decay and drop the arousal level to the resting state; if trials follow each other rapidly, the arousal level and generalization responsiveness level should tend to rise. Thus, arousal level and generalization would be interacting so as to retard extinction (or even overcome the effects of

[1] This study was carried out while the senior author was at Harvard University and was supported by grants from the National Association for Mental Health and National Science Foundation (Grant No. 3855). Elizabeth Brunette helped carry out a similar preliminary study using the dimension of pitch generalization. Her help with the apparatus of this study is greatly appreciated.

extinction and cause increasing responsiveness for some limited time).

Miller (1951) has hinted at the possibility of such a "bootstraps" phenomenon in anxiety. He questioned whether (in accordance with the results of studies using instrumental responses) massed trials would lead to the extinction of a learned fear response more rapidly than spaced trials. He also raised the possibility "that other effects, such as perseveration or summation of fear might reverse the (extinction) curve somewhat . . ." (p. 450).

In this study, it is predicted that the rate of extinction of a learned fear response and generalized responses will vary inversely with the speed of intertrial interval during extinction.

Method

The *S*s were trained to evidence a PGR reaction to a crucial word and tested with words which elicited this crucial word as an associate. In a factorial design, two intertrial intervals were used in conditioning and extinction.

Subjects.—The *S*s were 60 paid volunteer Harvard Summer School students. In order of appearance, they were assigned to four groups of 15 *S*s each, with the restriction that each subgroup was approximately evenly divided as to sex. The *S*s were informed that they would receive electric shocks.

Apparatus.—Galvanic measures were made with an apparatus suggested and described by Haggard and Gerbrands (1947). Stoelting finger electrodes were used with Cambridge electrode jelly.

The verbal stimulus materials were presented to *S* via earphones by means of a tape recorder. A measure of the intensity of the verbal stimuli revealed no differences between stimuli. A Harvard inductorium equipped with 1½-v. batteries delivered a shock which was judged quite painful but bearable by preliminary *S*s. This shock level was maintained for all *S*s, the batteries being periodically replenished.

Verbal stimulus materials.—The CS and mediated generalization stimuli (GS) were chosen by a method used by Martha Mednick (1957). They fit a paradigm of mediated generalization in which a CR is first trained to a CS. When a GS occurs which elicits the CS, the CR is in turn elicited. In this study the GS-CS connection was taken from the Russell and Jenkins (1954) revision of the norms of the Kent-Rosanoff Word Association Test. The auditory presentation of the word LIGHT was chosen as the CS. The GSs were the words LAMP and SOFT, both of which elicit LIGHT as a response (with probabilities of .63 and .0087, respectively). The word SQUARE was chosen as a neutral stimulus; it did not elicit LIGHT as a response in the revised norms. Thus, it was assumed that, after conditioning a PGR to LIGHT, when LAMP was presented it would elicit the implicit association LIGHT which in turn would be the stimulus for the conditioned PGR. It has been shown that this response generalizes from the auditory presentation to the implicit thought response (Noble, 1950).

During conditioning *S* heard, via earphones, seven neutral words and seven presentations of the CS. One-half second after the onset of five of the CS presentations *S* received a shock to the left wrist. Partial reinforcement during conditioning was used to prolong extinction and provide tests of conditioning level; the third and seventh presentations of the CS were not followed by the UCS.

The extinction series consisted of the CS, two generalization stimuli, and the neutral stimulus. These were presented four times each for a total of 16 extinction trials. The order of presentation in the extinction series was counterbalanced to avoid order effects.

Conditioning and extinction.—There were four groups. Group 8-20 had an 8-sec. intertrial interval in conditioning and a 20-sec. intertrial interval during extinction. These intertrial intervals were measured from the onset of CS in one trial to the onset of CS in the next trial. The other groups were Groups 8-8, 20-8, and 20-20.

The experiment was conducted in a quiet, dry, and cool basement room. The earphones limited the possibility of distracting noises. The *S* was shown the apparatus, seated in a broad-armed chair facing a wall, and had electrodes and earphones attached. It was explained that he would receive some electric shocks and hear some words; he was asked to remain quiet since physiological reactions were being taken. After an adaptation period in which *S* reached a steady PGR baseline, the conditioning series was begun. The generalization-extinction series followed the conditioning series with no interruption.

Results

Raw ohms resistance data were obtained from the records by measuring the heights of the momentary individual galvanic deflections and referring to a previously prepared conversion chart (Haggard & Gerbrands, 1947). Ohms resistance (r) was then transformed into micromhos (M) by use of the formula $M = 10^6 \frac{1}{r}$. Due to the skewed nature of the data, medians and appropriate nonparametric statistical analyses were used in almost all instances.

Conditioning intertrial interval.—The two CS trials not followed by shock during conditioning were included as a measure of extent of conditioning. However, for the groups having a conditioning intertrial interval of 8 sec. (8-8, 8-20), only the second of these trials was scoreable. The 8-sec. interval was not sufficient for *S* to recover from his UCR to the last UCS presentation. However, the one trial which was almost always scoreable was the unreinforced presentation of the CS at the conclusion of conditioning. This is an indicator of final conditioning level reached. The median response of the 20-20 and 20-8 groups was .0431 M; the median response of the 8-20 and 8-8 groups was .0179 M. The difference evaluated by the Mann-Whitney *U* test was significant ($z = 4.27$, $P < .001$), indicating that the 20-sec. intertrial interval in conditioning resulted in greater responsiveness to the CS at the end of conditioning.

Another way of evaluating the effect of the conditioning intertrial interval is to compare the resultant level of generalization responsiveness and ease of extinction. There were 16 generalization-extinction trials composed of the CS, two GSs, and the neutral stimulus, each presented four

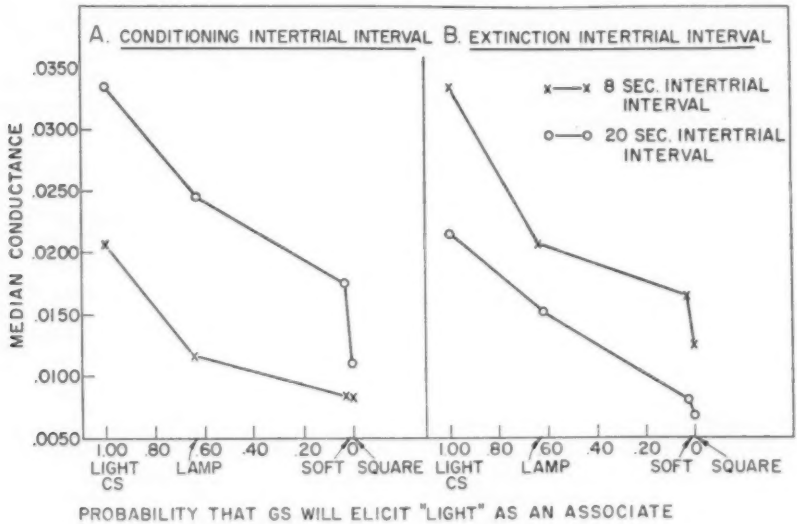

FIG. 1. Effect of intertrial interval during conditioning (A) and extinction (B) on generalization responsiveness.

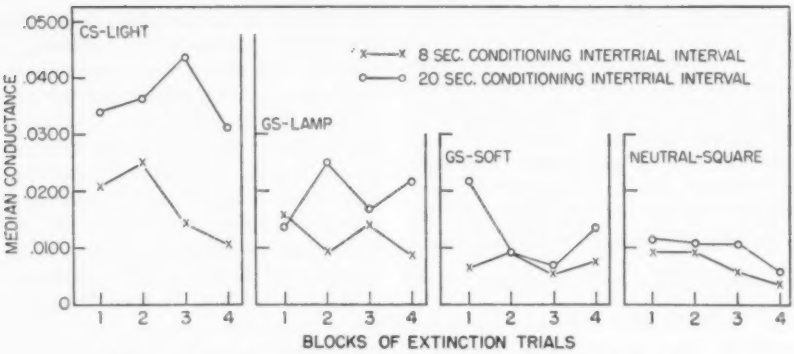

FIG. 2. Extinction of PGR to CS and GSs as a function of intertrial interval during conditioning.

times. By taking the median response to each GS and arranging the GSs in order of the probability that they would elicit the CS as an associated response, gradients of mediated generalization may be constructed for each group as in Fig. 1A. Figure 1A shows the effect of the intertrial interval during conditioning on generalization responsiveness during extinction. As can be observed, the 20-sec. intertrial interval in conditioning produced more extinction-generalization responsiveness than the 8-sec. conditioning intertrial interval. In order to obtain a total extinction-generalization score the amplitudes of responses to the test words were summed for each *S*. The differences between the 8-sec. and 20-sec. conditioning intertrial interval on this score proved significant in a Mann-Whitney *U* test ($z = 2.91$, $P < .001$).

Figure 2 compares the effect of the 20-sec. and 8-sec. intertrial intervals during conditioning on responsiveness in extinction. Extinction curves are plotted separately for the CS and GSs. The 20-sec. intertrial interval in conditioning resulted in slower

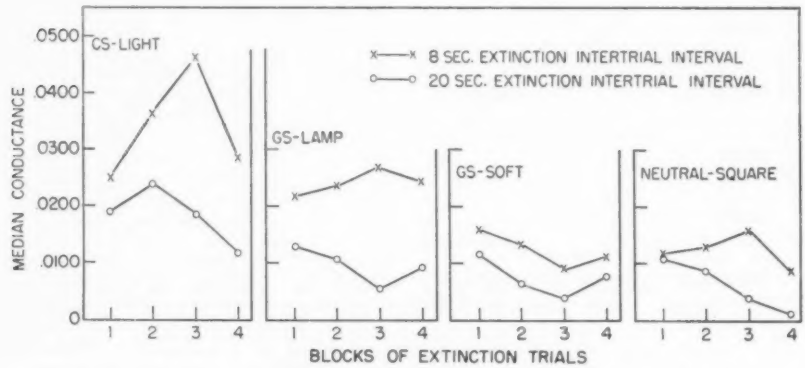

FIG. 3. Extinction of PGR to CS and GSs as a function of intertrial interval during extinction.

extinction of responsiveness to all of the stimuli.

Extinction intertrial interval.—Figure 1B shows the effect of the intertrial interval during extinction on generalization responsiveness. The 8-sec. intertrial interval during extinction resulted in more generalization responsiveness than the 20-sec. intertrial interval. The difference in total generalization score evaluated by the Mann-Whitney U test proved highly significant ($z = 3.21$, $P < .001$).

Figure 3 compares the effect of the 20-sec. and 8-sec. intertrial intervals during extinction on responsiveness during extinction. Extinction curves are plotted separately for the CS and GSs. The 8-sec. intertrial interval in extinction resulted in slower extinction than the 20-sec. extinction intertrial interval. For the CS and GS, LAMP, there is no evidence of any lessening of responsiveness at the conclusion of the extinction series; if anything, the amplitude of responses seems to increase as "extinction" proceeds.

Inspection of the extinction curves for the four subgroups did not suggest that there had been any interactive effect of conditioning and extinction intertrial intervals. The curves for Groups 20-8 and 8-8 were essentially parallel as were the curves of Groups 8-20 and 20-20. An analysis of variance supported this impression; the interaction term was not significant ($F = .12$, $df = 1/56$). The main effects were significant.

Discussion

Spaced conditioning trials resulted in greater amplitude of response at the conclusion of conditioning, more generalization, and more resistance to extinction than massed conditioning trials. While this finding is in accord with most previous research and theory (Hull, 1943; Rohrer, 1947) it is somewhat surprising in this study. The effectiveness of the distributed practice had to be demonstrated in the face of the presumably higher arousal level under the 8-sec. conditioning intertrial interval. However, the conditioning amplitude result should be interpreted with caution since it is based on data from only one trial.

Massed extinction trials with the CS and GSs resulted in greater generalization and more resistance to extinction than spaced extinction trials. A similar result for resistance to extinction has been reported by Sheffield (1950). She suggested that the massed trials group suffered greater frustration in extinction leading to greater drive leading to greater resistance to extinction. The results of this study seem to extend the generality of her interpretation. That is, any aspect of the extinction procedure that will increase arousal will tend to slow extinction. The more arousal provoked, the slower the extinction process. In the present study special advantage was taken of the arousal properties of the fear response. In the massed trials condition, fear eliciting stimulation (CS and GSs) was administered before the fear from the preceding trial could decay, making for stronger fear responses, greater accumulation of arousal, and slower extinction. These results are even more impressive in view of the fact that extinction of response to generalization stimuli is usually very rapid. Two recent studies have reported similar findings. Both used a noxious UCS and found more resistance to extinction with massed trials in extinction; both human and animal *S*s were studied. Murphy and Miller (1956) used an electric shock UCS in a modified Warner box; Howat and Grant (1958) studied eyelid conditioning. Both found faster extinction under spaced extinction trials.

The relationship between the reciprocal augmentation hypothesis and the acute schizophrenic process is discussed elsewhere (Mednick, 1958).

Summary

A psychogalvanic reaction was trained to the word LIGHT under conditions of massed and distributed trials. In a factorial design the CR to the CS, LIGHT and mediated generalization stimuli, LAMP, SOFT, and SQUARE, were extinguished with 8-sec. and 20-sec. intertrial intervals.

Spaced trials in conditioning produced greater amplitude of CR, greater generalization, and more resistance to extinction. Massed trials in extinction led to greater generalization and slower extinction. An interpretation of the results is offered in terms of reciprocal augmentation of anxiety and generalization, a phenomenon thought to have relevance for psychopathology.

References

HAGGARD, E. A., & GERBRANDS, R. An apparatus for the measurement of continuous changes in palmar skin resistance. *J. exp. Psychol.*, 1947, **37,** 92–98.

HOWAT, M. G., & GRANT, D. A. Influence of intertrial interval during extinction on spontaneous recovery of conditioned eyelid responses. *J. exp. Psychol.*, 1958, **56,** 11–15.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, **64,** 276–287.

MEDNICK, M. T. Mediated generalization and the incubation effect as a function of manifest anxiety. *J. abnorm. soc. Psychol.*, 1957, **55,** 315–321.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychol. Bull.*, 1958, **55,** 316–327.

MEDNICK, S. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57,** 169–200.

MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

MURPHY, J. V., & MILLER, R. E. Spaced and massed practice with a methodological consideration of avoidance conditioning. *J. exp. Psychol.*, 1956, **52,** 77–81.

NOBLE, C. E. Conditioned generalization of the GSR to a subvocal stimulus. *J. exp. Psychol.*, 1950, **40,** 15–25.

ROHRER, J. H. Experimental extinction as a function of the distribution of extinction trials and response strength. *J. exp. Psychol.*, 1947, **37,** 473–493.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *Off. Naval Res. tech. Rep.*, 1954, No. 11. (Contract No. 8-ONR-66216)

SHEFFIELD, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40,** 305–313.

(Received May 18, 1961)

Deterrents and Reinforcement

THE PSYCHOLOGY OF INSUFFICIENT REWARD

Douglas H. Lawrence and Leon Festinger

Festinger's theory of dissonance is applied to the problem of why partial reward, delay of reward, and effort expenditure during training result in increased resistance to extinction. The authors present a critical review and new data on alternative explanations, concluding with a look at the broader implications of the new theory. $4.75

A Theory of Cognitive Dissonance

Leon Festinger

A reissue and new printing of a book first published in 1957 and hailed as one of the seminal contributions to modern psychological theory. "An idea of considerable power and originality."—*The American Journal of Psychology*. "A major theoretical statement on the general topic of communication and social influence. . . . An impressive step in formalizing theory in the area."—*Annual Review of Psychology* (1959). $5.25

Personality and Decision Processes

STUDIES IN THE SOCIAL PSYCHOLOGY OF THINKING

Orville G. Brim, Jr., David C. Glass
David E. Lavin, and Norman Goodman

The development of new concepts and generalizations about the relations between decision making, personality, and social structure is the focus of these studies, which have been designed and carried out from the point of view of the sociologist. Information was obtained from 200 adults through personality tests and tests of decision-making processes. About $8.00

Order from your bookstore, please

STANFORD UNIVERSITY PRESS

Wiley BOOKS

INTRODUCTION to PSYCHOLOGICAL RESEARCH

By William A. Scott *and* Michael Wertheimer, *both of the University of Colorado.* Any researcher encounters new and unforeseen difficulties in each project that he undertakes. The seasoned investigator is equipped to think critically and creatively about these problems because he has mastered the basic techniques for handling the difficulties that crop up in every project. Some problems appear constantly, in varying forms, in most research projects. Mastery of techniques for handling these problems will free the investigator's attention for the more specific problems inherent in his chosen project. But this mastery must be based upon a comprehension of the principles and reasons behind the mechanical techniques. If the investigator fully understands these basics, not only can he handle the recurrent problems, but he is equipped to invent new techniques appropriate to his research interest.

The recurrent problems include questions such as how to formulate research questions for empirical study, how to find out what has been done in the research area, should an experiment or less formal observations be attempted, what operational definitions should be used, how many subjects should be obtained and what kinds of statistics are required, and how to write up the findings and find an appropriate publishing outlet for them.

This book focuses upon these kinds of issues with the aim of showing a pattern to the research process which the researcher can apply to his own research experiments. The authors discuss various methodologies and research techniques which will be helpful in many experiments. This is not an intensive treatment of specialized topics but an introduction to research methods. The authors base their approach upon an ideological concept of science which includes a desire to nurture the creative spark by an emphasis upon flexibility within the limits of producing convincing evidence. They intend to aid the creative process by introducing the reader to techniques which will free him to concentrate more fully upon the essential research problem. This book is a down-to-earth introduction to the practicalities of doing original behavioral research. 1962. 445 pages. $7.95.

DETERMINANTS OF INFANT BEHAVIOR

Edited by B. M. Foss, University of London. 1961. 307 pages. $6.75.

FRUSTRATION AND CONFLICT

By A. J. Yates, University of New England. 1962. In press.

Send for examination copies

JOHN WILEY & SONS, Inc.

440 Park Avenue South **New York 16, N.Y.**